

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:18:30 ; Search time 84.6574 Seconds
(without alignments)
1632.058 Million cell updates/sec

Title: US-09-455-978B-2

Perfect score: 2394

Sequence: 1 MSNDNDTLVTADVENGIDGH.....ATDQQRVTEVEVRETVGKLS 489

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04.*

- 1: geneseqp1980s.*
- 2: geneseqp1990s.*
- 3: geneseqp2000s.*
- 4: geneseqp2001s.*
- 5: geneseqp2002s.*
- 6: geneseqp2003as.*
- 7: geneseqp2003bs.*
- 8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2394	100.0	489	AAE04641	Aae04641 Halobacte
2	412.5	17.2	432	AAE04642	Aae04642 Bacillus
3	399.5	16.7	535	ABU38469	Abu38469 Protein e
4	387.5	16.2	739	AAE04643	Aab96493 Putative
5	386.5	16.1	682	ABU38250	Abu38250 Protein e
6	380	15.9	435	AAE04648	Aab96483 Putative
7	375	15.7	501	AAE046709	Abu96709 Putative
8	374.5	15.6	680	ABU41746	Abu41746 Protein e
9	373.5	15.6	709	ADA33885	Ada33885 Acinetoba
10	371.5	15.5	686	ABU40238	Abu40238 Protein e
11	361	15.1	644	ABU49667	Abu49667 Protein e
12	359.5	15.0	891	AAE26596	Aab26596 Synchocy
13	356.5	14.9	374	AAE046382	Aab96382 Putative
14	356.5	14.9	539	ABU23101	Abu23101 Protein e
15	348	14.5	626	ABU49142	Abu49142 Protein e
16	346.5	14.5	565	ABM68096	ABM68096 Photorhab
17	339.5	14.2	601	ABE49909	ABE49909 Listeria
18	338	14.1	686	ABU24381	Abu24381 Protein e
19	326	13.6	545	ABM68095	ABM68095 Photorhab
20	326	13.6	641	ABU49354	Abu49354 Protein e
21	315	13.2	643	ABU49678	Abu49678 Protein e
22	314.5	13.1	501	ABU41640	Abu41640 Protein e
23	313.5	13.1	541	ABU49123	Abu49123 Protein e
24	308	12.9	845	ABU48593	ABU48593 Protein e
25	298	12.4	511	AAE14148	AAE14148 Bordetell

ALIGNMENTS

RESULT 1
AAE04641

ID AAE04641 standard; protein; 489 AA.

XX AAE04641;

XX DT 11-SEP-2003 (revised)

XX DT 04-SEP-2001 (first entry)

XX DE Halobacterium salinarum HemAT-Hs protein.

XX KW Haem binding protein; HemAT-Hs; HemAT-Bs; gaseous ligand sensor;
XX KW oxygen storage; artificial photosynthesis; signalling function;
XX KW alpha-haemoglobin; myoglobin; therapy.

XX OS Halobacterium salinarum.

XX EN WO200140475-A2.

XX PD 07-JUN-2001.

XX PF 05-DEC-2000; 2000WO-US033048.

XX PR 06-DEC-1999; 99US-00455978.

XX XX (UYHA-) UNIV HAWAII.

XX FI Alam M, Larsen R;

XX DR WPI; 2001-374832/39.

XX N-PSDB; AAD08991.

XX PT Novel isolated bacterial heme binding protein, HemAT-Hs or HemAT-Bs which
XX reversibly binds oxygen with low affinity, useful for controlled storage
XX of oxygen and for sensing gaseous ligands such as oxygen.

XX PS Claim 6; Page 10; 94pp; English.

XX CC The present invention relates to isolated archeal and bacterial haem
XX binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen with
XX low affinity. Haem binding protein is useful for controlled storage of
XX oxygen by allowing haem binding protein to bind and store oxygen, and
XX triggering the release of oxygen from haem binding protein by activating
XX the signalling domain. Haem binding protein is useful for sensing gaseous
XX ligands such as oxygen, NO, CO, or CN. Blood substitute comprising low blood
XX binding protein is useful for treating a patient suffering from haem-
XX levels by administering and regulating the oxygen binding of the haem-
XX binding protein by modifying the signalling domain. Haem binding protein

Abg24223 Novel hbm
Aaw98798 H. pylori
Aaw71554 Helicobac
Aab46327 H. pylori
Aaw98797 H. pylori
Aaw71553 Helicobac
Aab46345 H. pylori
Aab10361 H. pylori
Abu51211 Helicobac
Aaw20287 H. pylori
Aaw20769 H. pylori
Aay19897 B. burgdo
Aay19896 B. burgdo
Abu19306 Protein e
Aay19800 B. burgdo
Aay19799 B. burgdo
Aab52596 Helicobac
Abu50773 Helicobac
Aaw05196 Helicobac
Abb49349 Listeria

26 297.5 12.4 1137 4 ABG24223
27 289 12.1 565 2 AAW98798
28 289 12.1 565 2 AAW71554
29 289 12.1 565 4 AAB46327
30 284 11.9 630 2 AAW98797
31 284 11.9 630 2 AAW71553
32 284 11.9 675 4 AAB46345
33 280 11.7 675 2 AAY10361
34 277.5 11.6 289 5 ABUS1211
35 276 11.5 293 2 AAW20287
36 273.5 11.4 664 2 AAW20769
37 269 11.2 723 2 AAY19897
38 269 11.2 753 4 AAY19896
39 269 11.2 753 6 ABU19306
40 266 11.1 606 2 AAY19800
41 266 11.1 633 2 AAY19799
42 257 10.7 431 3 AAB52596
43 257 10.7 431 5 ABUS0773
44 254 10.6 433 2 AAW05196
45 253 10.6 300 5 ABB49349

CC is useful for haem-based catalysis, for artificial photosynthesis and for
 CC identifying potential signalling functions of mutated alpha-haemoglobin
 CC and myoglobin causing several diseases. The present sequence is
 CC Halobacterium salinarium HemAT-Hs protein which is salt tolerant.
 CC (Updated on 11-SEP-2003 to standardise OS field)
 XX
 SQ Sequence 489 AA;

Query Match 100.0%; Score 2394; DB 4; Length 489;
 Best Local Similarity 100.0%; Pred. No. 6.3e-165;
 Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNDNDTLVTADVRNGIDGHALADRIGLDEARIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
 DB 1 MSNDNDTLVTADVRNGIDGHALADRIGLDEARIAWRLSFTGIDDDTMAALAAEQPLFEAT 60

QY 61 ADALVTDFYDHLSEYRTQDLFANSTKTKVEQLKETOABYLLGLGRGEYDTEYAAQRARIG 120
 DB 61 ADALVTDFYDHLSEYRTQDLFANSTKTKVEQLKETOABYLLGLGRGEYDTEYAAQRARIG 120

QY 121 KIHVDVLGPGDVLGAYTRYTYTGLLDALADDVADRGHEMAAANDVAVLPMLKLTFF 180
 DB 121 KIHVDVLGPGDVLGAYTRYTYTGLLDALADDVADRGHEMAAANDVAVLPMLKLTFF 180

QY 181 DQOIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDMRART 240
 DB 181 DQOIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDMRART 240

QY 241 DQQVDRMADVSGREISSVSAEVEASTADDVRRTSSEDAEALAAQGEAAADDAATMTDID 300
 DB 241 DQQVDRMADVSGREISSVSAEVEASTADDVRRTSSEDAEALAAQGEAAADDAATMTDID 300

QY 301 EATDQVTAAGVQLGGERAADVSVTVGIDDIABOTNMALNASIEAARAGEGEGFAVVAD 360
 DB 301 EATDQVTAAGVQLGGERAADVSVTVGIDDIABOTNMALNASIEAARAGEGEGFAVVAD 360

QY 361 EVKALAEBSREGSTVEBELVEQMAETETVDQLDEVNQRIGEGVERVEEAMETLQETID 420
 DB 361 EVKALAEBSREGSTVEBELVEQMAETETVDQLDEVNQRIGEGVERVEEAMETLQETID 420

QY 421 AVEDAASGMQVSTATDQAVSTERVAEWDGVDVDRAGEIAAALDDIADATDQQVTVTEE 480
 DB 421 AVEDAASGMQVSTATDQAVSTERVAEWDGVDVDRAGEIAAALDDIADATDQQVTVTEE 480

QY 481 VRETGKLS 489
 DB 481 VRETGKLS 489

RESULT 2
 AAE04642
 ID AAE04642 standard; protein; 432 AA.

XX AC
 XX AAE04642;

XX DT
 XX 04-SEP-2001 (first entry)

XX DE
 XX Bacillus subtilis HemAT-Bs protein.

XX KW
 XX Haem binding protein; HemAT-Hs; HemAT-Bs; gaseous ligand sensor;
 XX oxygen storage; artificial photosynthesis; signalling function;
 XX alpha-haemoglobin; myoglobin; therapy.

XX OS
 XX Bacillus subtilis.

XX PN
 XX WO200140475-A2.

XX PD
 XX 07-JUN-2001.

XX PF
 XX 05-DEC-2000; 2000WO-US033048.

XX PR
 XX 06-DEC-1999; 99US-00455978.

XX

PA (UYHA-) UNIV HAWAII.
 XX
 PI Alam M, Larsen R;
 XX
 DR WPI; 2001-374832/39.
 DR N-PSDB; AAD08992.
 XX

PT Novel isolated bacterial heme binding protein, HemAT-Hs or HemAT-Bs which
 PT reversibly binds oxygen with low affinity, useful for controlled storage
 PT of oxygen and for sensing gaseous ligands such as oxygen.

XX
 PS Claim 8; Page 11; 94pp; English.

XX The present invention relates to isolated archeal and bacterial haem
 CC binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen with
 CC low affinity. Haem binding protein is useful for controlled storage of
 CC oxygen by allowing haem binding protein to bind and store oxygen, and
 CC triggering the release of oxygen from haem binding protein by activating
 CC the signalling domain. Haem binding protein is useful for sensing gaseous
 CC ligands such as oxygen, NO, CO, or CN. Blood substitute comprising haem
 CC binding protein is useful for treating a patient suffering from low blood
 CC levels by administering and regulating the oxygen binding of the haem-
 CC binding protein by modifying the signalling domain. Haem binding protein
 CC is useful for haem-based catalysis, for artificial photosynthesis and for
 CC identifying potential signalling functions of mutated alpha-haemoglobin
 CC and myoglobin causing several diseases. The present sequence is Bacillus
 CC subtilis HemAT-Bs protein

XX SQ Sequence 432 AA;

Query Match 17.2%; Score 412.5; DB 4; Length 432;
 Best Local Similarity 26.2%; Pred. No. 1.4e-21;
 Matches 114; Conservative 85; Mismatches 187; Indels 49; Gaps 8;

QY 31 AEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFYDHLSEYRTQDLFANSTKTV 90
 DB 32 ADVKQLKMWRLGDAEYLVLEQLQENIVNIDAFYKNLDHESLMDII-NDHSSVD 90

QY 91 QLKETOABYLLGLGRGEYDTEYAAQRARIGKIHVDVLGPGDVLGAYTRYTYTGLLDALAD 150
 DB 91 RLKQTLKRIHQEMFAGVIDDEFIEKRNRIASIHRLRIGLLPKWYMGAFQELLMSMDIY-- 148

QY 151 DWADRGEEAAAANDVAVLPMLKLTTFDQQLAMDTYIDSYAQRLHDEIDSRQELANA 210
 DB 149 -----EASITNQOELLKAIKATTKILNLGQVLLEAFQSEYNQ-TRDEQEEKNLH- 199

QY 211 VATHVEAPLSSLEATSQDVAERTDMRARTDQVDRMADVSGREISSVSAEVEASTADD 270
 DB 200 -----QKIQTSGSIA-----NLFSETSRVQELVDKSEGISQASKAGTVTSSTVEE 246

QY 271 VRTSSEDAEALAAQGEAAADDAATMTDIDATDGTAGVQLGERAADVSVTVGIDDI 330
 DB 247 -KSIGGKKELBVQKQ-----MNKIDTSLVQIEKEMVKLDEIAQOIKIFGIVTGI 296

QY 331 AEQTNMLALNASIEAARAGEGFAVVADVEKALAEBSREGSTVEBELVEQMAETET 390
 DB 297 AEQTNLLSLNASIESARAGEHGKGFVAVNEVRKLSEOTKKTSTVSELVNTNTQINIV 356

QY 391 VDQLDEVNQRIGEGVERVEEAMETLQETIDDAVEDA-----ASGMQVSTAT 436
 DB 357 SKHINDVNLVSEKSKMTQINRLFDTEIVHSMKISKEQSGKIDVDLQAFGLGQVSRVAV 416

QY 437 DEQAVSTEEVAEMVD 451
 DB 417 SHVAASVDSLILTE 431

RESULT 3

ABU38469
 ID ABU38469 standard; protein; 535 AA.

XX AC
 XX ABU38469;

XX

19-JUN-2003 (first entry)
 Protein encoded by Prokaryotic essential gene #23996.
 Antisense; prokaryotic essential gene; cell proliferation; drug design.
 Pseudomonas aeruginosa.
 WO200277183-A2.
 03-OCT-2002.
 21-MAR-2002; 2002WO-US009107.
 21-MAR-2001; 2001US-00815242.
 06-SEP-2001; 2001US-00948993.
 25-OCT-2001; 2001US-0342923P.
 08-FEB-2002; 2002US-00072851.
 06-MAR-2002; 2002US-0362699P.
 (ELIT-) ELITRA PHARM INC.
 Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,
 Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 WPI; 2003-029926/02.
 N-PSDB; ACA42339.
 New antisense nucleic acids, useful for identifying proteins or screening
 for homologous nucleic acids required for cellular proliferation to
 isolate candidate molecules for rational drug discovery programs.
 Claim 25; SEQ ID NO 66393; 1766pp; English.
 The invention relates to an isolated nucleic acid comprising any one of
 the 6213 antisense sequences given in the specification where expression
 of the nucleic acid inhibits proliferation of a cell. Also included are:
 (1) a vector comprising a promoter operably linked to the nucleic acid
 encoding a polypeptide whose expression is inhibited by the antisense
 nucleic acid; (2) a host cell containing the vector; (3) an isolated
 polypeptide or its fragment whose expression is inhibited by the
 antisense nucleic acid; (4) an antibody capable of specifically binding
 the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 proliferation or the activity of a gene in an operon required for
 proliferation; (7) identifying a compound that influences the activity of
 the gene product or that has an activity against a biological pathway
 required for proliferation, or that inhibits cellular proliferation; (8)
 identifying a gene required for cellular proliferation or the biological
 pathway in which a proliferation-required gene or its gene product lies
 or a gene on which the test compound that inhibits proliferation of an
 organism acts; (9) manufacturing an antibiotic; (10) profiling a
 compound's activity; (11) a culture comprising strains in which the gene
 product is overexpressed or underexpressed; (12) determining the extent
 to which each of the strains is present in a culture or collection of
 strains; or (13) identifying the target of a compound that inhibits the
 proliferation of an organism. The antisense nucleic acids are useful for
 identifying proteins or screening for homologous nucleic acids required
 for cellular proliferation to isolate candidate molecules for rational
 drug discovery programs, or for screening homologous nucleic acids
 required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
K. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of
 the target prokaryotic essential genes. Note: The sequence data for this
 patent did not form part of the printed specification, but was obtained
 in electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences
 Sequence 535 AA;
 Query Match 16.7%; Score 399.5; DB 6; Length 535;
 Best Local Similarity 25.2%; Pred. No. 1.6e-20;
 Matches 132; Conservative 111; Mismatches 206; Indels 75; Gaps 16;
 3 NDNDTLTVADVRNGIDGHALADRLGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATAD 62

Db 37 SENELSVNA-LRNHEGDMHED-----ALRADVLAFFV--OPGDGAARE 78
 Qy 63 ALVTDFDYHLESYERTQDLFANSTKTVEQ-----LKTQAEYLLGLGRGYDTEYAAQRA 117
 Db 79 QVRQDLQEHSGWFR-----KVVEQNGQLFLNDAIHQALVEL-RPDLEAYIGAAS 127
 Qy 118 RIGK-IHDVLGLGPDV--YLGATRYTYTGLLDALADDV-----VADRGEE----- 159
 Db 128 IVGKALLDPVAARAELPFQVQAF-KBLEGRNEALSSLIETKHEVQTNREARSDMRYSAWML 186
 Qy 160 -----AAAAVDELVARFL-----PMLKLLTFDDQIAMDTY-----IDS--YAORLHDEI 201
 Db 187 AGGILVACLVLGQLCRQLLRAVLQPLKLVSSARVIAQGNLOEPIGVDSNDEAQLQAL 246
 Qy 202 DSRQELANAVATHVEAPLSSLEATSQDAERTDTMRARTDDQVDRMADVSREISVSASV 261
 Db 247 GEMQENLRMITIIRQSESEELHDTQSIGQTSQSIHVHGASQQAADSATSMAASMEEMITNI 306
 Qy 262 EVASTADDVVRTSEDAEALAOQGEAAADALATMTDIDEATDGVTAGVEOLGERAADVE 321
 Db 307 SQISDHADNARVISAKSEELASSGGQVILNVVEGMSRIADVNVQSSTSI TALGQSSDEIH 366
 Qy 322 SVTGVIDDIAEOTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREOSTRVEELVE 381
 Db 367 SIIQVIKIGIAEQTNLLALNAIEAARAGEAGEGFAVVADEVKALAEESREOSTRVEELVE 426
 Qy 382 QMQAETETVDQDEVNORIGEGVERVEEAMETLQETDAVEDAASGMQEVSTATDEQAV 441
 Db 427 RIRASTQGAINSMEAGVSRVNEGVSFARQAGVSNELIDGTRHAASVVDEISQITIREQSR 486
 Qy 442 STEVAEMVGVDDVRAGEIAAALDDIADATDQQVTVVEEVRETIV 485
 Db 487 ASDEIAQRVELIAQRSQONTQAMHEMA-AT---ARRLNEVAATM 526
 RESULT 4
 AAB96493
 ID AAB96493 standard; protein; 739 AA.
 AC AAB96493;
 XX 29-OCT-2001 (first entry)
 DE Putative sensory transduction histidine kinase and response regulator #3.
 XX Hyperthermophilic archaeon; hyperthermophilic protein.
 XX Pyrococcus abyssi.
 XX FR2792651-A1.
 XX 27-OCT-2000.
 XX 21-APR-1999; 99FR-00005034.
 XX 21-APR-1999; 99FR-00005034.
 XX (CNRS) CNRS CENT NAT RECH SCI.
 XX (IPRE-) IPREMER INST FR RECH EXPL MER.
 PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 PI Querellou J, Weissenbach J, Saurin W, Heilig R;
 XX WPI; 2001-126236/14.
 XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins
 useful in industry.
 XX Claim 7; Page 1203-1205; 1657pp; French.
 XX The present invention relates to the genomic sequence of *Pyrococcus*
 CC abyssi (see AAF96431 and AAH41223-7) and *P. abyssi* proteins. *P. abyssi* is

Db 626 SSAGHSNTMNVQIBTSQTSAGTATARSIGNLAKMASEMNSV 670

RESULT 6

AAB96483

ID AAB96483 standard; protein; 435 AA.

XX AC AAB96483;

XX XX

DT 29-OCT-2001 (first entry)

XX XX

DE Putative sensory transduction histidine kinase and response regulator #2.

XX XX

KW Hyperthermophilic archaeon; hyperthermophilic protein.

XX XX

OS Pyrococcus abyssi.

XX XX

FN FR2792651-A1.

XX XX

PD 27-OCT-2000.

XX XX

PF 21-APR-1999; 99FR-00005034.

XX XX

PR 21-APR-1999; 99FR-00005034.

XX XX

(CNRS) CNRS CENT NAT RECH SCI.

PA (IFRE-) IPREMER INST FR RECH EXPL MER.

XX XX

PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

XX XX

PI Querellou J, Weissensbach J, Saurin W, Heilig R;

XX XX

DR WPI; 2001-126236/14.

XX XX

XX New nucleotide sequences isolated from Pyrococcus abyssi encode proteins

PT useful in industry.

XX XX

PS Claim 7; Page 1189-1190; 1657pp; French.

XX XX

CC The present invention relates to the genomic sequence of Pyrococcus

CC abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is

CC a hyperthermophilic archaeon, which is isolated from deep-sea

CC hydrothermal vents. The present sequence is one such P. abyssi protein.

CC The proteins of the present invention have various potential industrial

CC uses, since the proteins are stable at very high temperatures, some up to

CC 110 degrees centigrade. Note: This patent is in the same patent family as

CC WO200065062, which contains additional sequences as shown in AAB99132-

CC AAB99143, AAH75903-AAH75920 and AAG66436

XX XX

SQ Sequence 435 AA;

XX XX

Query Match

Best Local Similarity

Matches 109; Conservative

91; Mismatches 129; Indels 68; Gaps 12;

QY 107 EYDEYAQRARIGIKHIVLGLGPDVYLGA

DB 63 EYEKQLOREIDEIVKLDRIAQG-DL-----S

QY 167 LVARFLPMLKLLTDFDQOIAMDTYIDSYAQR

DB 108 L-RKSIQVLNINAKAIDVRNHTKVKMENTEQ

QY 226 SQDVAERTDTHRAITDQVDRMADVRSISSV

DB 162 QENISKMTDTR-----YIHDSKE--TVSTW

QY 286 EAAADDAATWTDIDEATDGTAGVQOLGERA

DB 208 ROAADQ-IEELSRWMEKIEETVRGAENGK--

QY 346 ARAGEGEGFVAVADEVKALAESEREQSTRV

DB 107 EYDEYAQRARIGIKHIVLGLGPDVYLGA

DB 63 EYEKQLOREIDEIVKLDRIAQG-DL-----S

QY 167 LVARFLPMLKLLTDFDQOIAMDTYIDSYAQR

DB 108 L-RKSIQVLNINAKAIDVRNHTKVKMENTEQ

QY 226 SQDVAERTDTHRAITDQVDRMADVRSISSV

Db 264 ARAGEAGKGFPAVDAETKRLAESKKAADIRELIQ-----IGDKIGESV 309

QY 406 ERVEEAMETLQEIITDAVEDAASGMQEVSTATDQAVSTEEVAEMV-----D 451

Db 310 EVTQOGAEVVKTSDEVIKESVSYLTQVAEMMEVEVKASELREKVIQEGKIEEGLRFL 369

QY 452 GVDDRAGEIAAALDDIADATDQOVRTVEEVRETVGKL 488

Db 370 NLAASAEETTAABEVSAABEQSSALQELRESVKEL 406

RESULT 7

AAB96709

ID AAB96709 standard; protein; 501 AA.

XX AC AAB96709;

XX XX

DT 29-OCT-2001 (first entry)

XX XX

DE Putative sensory transduction histidine kinase & response regulator #4.

XX XX

KW Hyperthermophilic archaeon; hyperthermophilic protein.

XX XX

OS Pyrococcus abyssi.

XX XX

FN FR2792651-A1.

XX XX

PD 27-OCT-2000.

XX XX

PF 21-APR-1999; 99FR-00005034.

XX XX

PR 21-APR-1999; 99FR-00005034.

XX XX

(CNRS) CNRS CENT NAT RECH SCI.

PA (IFRE-) IPREMER INST FR RECH EXPL MER.

XX XX

PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

XX XX

PI Querellou J, Weissensbach J, Saurin W, Heilig R;

XX XX

DR WPI; 2001-126236/14.

XX XX

XX New nucleotide sequences isolated from Pyrococcus abyssi encode proteins

PT useful in industry.

XX XX

PS Claim 7; Page 1469-1471; 1657pp; French.

XX XX

CC The present invention relates to the genomic sequence of Pyrococcus

CC abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is

CC a hyperthermophilic archaeon, which is isolated from deep-sea

CC hydrothermal vents. The present sequence is one such P. abyssi protein.

CC The proteins of the present invention have various potential industrial

CC uses, since the proteins are stable at very high temperatures, some up to

CC 110 degrees centigrade. Note: This patent is in the same patent family as

CC WO200065062, which contains additional sequences as shown in AAB99132-

CC AAB99143, AAH75903-AAH75920 and AAG66436

XX XX

SQ Sequence 501 AA;

XX XX

Query Match

Best Local Similarity

Matches 116; Conservative

111; Mismatches 175; Indels 74; Gaps 12;

QY 65 VTFPYDHL-----SYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDYAA 114

DB 30 ISLDYDNLEKTMKVQVQGSVEAIRETLENAK---SQKET-----LWLSIGVMSI----- 77

QY 115 QRARIGIHVDVGLGPDVYLGA

DB 78 -----VAVISGALG-----RLMNSTMRPFINEMAKIAESIAEGKLSRAREMYSKIYREDDE 129

QY 175 L-KLLTFDQOIAMDTY-----IDSYAQR

DB 130 IGLKIEGRAISQDVLQTLLEIITRM--EKISG

QY 227

DB 187

Qy 228 DVARTDMR-----ARTDDQVRMDVSRREISSVSASVEEVA----- 265
 Db 198 NLRLMKTVRDLALTLEGRANDLTRISSEISEAINQVAEATQVSVSVEAQQOENITEIME 247
 Qy 266 ---STADVRRTSDEAE-----ALAOQGEAAADALATMTDIDEATDGTAGVEQ 312
 Db 248 GWNITADVTQRTVDAMEEFGVNVESIAREGKDKGKAISQVEDIQDAMKVIRQAOVE 307
 Qy 313 LGERAADVESVTGVDDIAEQTNMLNLSIAEAPAGEAGGFVAVDEVKALAPESREQ 372
 Db 308 VAEMSKNVGDIINALADIAEQTNMLNLSIAEAPAGEAGELGRFVAVQEVNLAESKEA 367
 Qy 373 STRVEELVEQMAETEEFTVDQDEVNQIRIGEVERVEEAMETLQBITDAVEDAASGMQEV 432
 Db 368 AEKIRGILNEIQEKVAKVEETKGVKVVDSVDFLKETVGLMNMIGELDDVESKLQDI 427
 Qy 433 STATDEQAVTEEVAEVMDVGDVDRAGETAAALDDIADATDQVRVTEVRETVGKL 488
 Db 428 KNELANQTEHVENAKALENLAASAQETTASAEVVSASQAQESMEVEVKRNIIEL 483

RESULT 8

ABU41746
 ID ABU41746 standard; protein; 680 AA.

AC ABU41746;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #27273.

DE Antisense; prokaryotic essential gene; cell proliferation; drug design.

KW Pseudomonas syringae.

OS WO200277183-A2.

PN 03-OCT-2002.

PD 21-MAR-2002; 2002WO-US009107.

PF 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR N-PSDB; ACA45616.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 69670; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 680 AA;

Query Match 15.6%; Score 374.5; DB 6; Length 680;

Best Local Similarity 24.2%; Pred. No. 1.4e-18;

Matches 120; Conservative 103; Mismatches 201; Indels 71; Gaps 12;

Qy 23 ADRIGLDEAEIA-----WRLSFTGIDD-DTMAALAAEQPLFE---ATADALVTDF 68

Db 207 ADAFGRDASQFGRVLNGLMGNATLIRITQVEDARARLARIAELFEVFSVSDILRTS 266

Qy 69 YDHLSEYERTODLFANSTKTVQELKETOAEYLLGLGRGEYDEYAAQARIGKIH-D-VLG 127

Db 267 PELYQVREASGNIF-NTSQTLDDTSVLANSI-----ENLAKRTWNTVGGYVLG 315

Qy 128 LGPDVYLGAITYTYTGLLDALADDVADRGEAAAVDELVARPLMLKLLTFDQOIA-- 185

Db 316 L-----LALMSILILGLV-----MVRETNRQLRETAQKSERNQTAIMRLDEIENLADG 364

Qy 186 -----MDYIIDSYAORLHDEIDSRQELA---NAVATHVEAPLSSLEATSDQVAERTD 234

Db 365 DLTVTASVTEDFTGAIDSNYSIDQLRELVTINLTAEQVASAVTETQATAMQLS---- 420

Qy 235 TMRARTDDQVRMDVSRREISSVSASVEEVASTADDVRRTSEDAAALAOQGEAAADALA 294

Db 421 ---AASEHQALQISAASTAVNDMAASIDQVSAASESSAVALERSVAIAKNGEVVQNTIH 477

Qy 295 TMTDIDEATDGTAGVEQLGERAADVESVTGVDDIAEQTNMLNLSIAEAPAGEAGEG 354

Db 478 GMDNIREQIQDTSKRIKLGESSQBEIGDIVSLDIDDIADQTNILALNAAIQASMGADGRG 537

Qy 355 FAVVADEVKALAESREQSTVEELVEQMOAETETVDQDEVNQIRIGEVERVEEAMET 414

Db 538 FAVVADEVORLAERSSSSATKQIETLIVRAIQNDTNEAVISMEQTTSSEVVGRGARLAQDAGVA 597

Qy 415 LQETIDAVEDAASGMQEVSTATDEQAVSTEEVAEVMDVGDVDRAGETAAALDDIADATDQ 474

Db 598 LQETEGSVRLAEILIESITDAHQQAES-----AGQISQTMVTIQTTSQT 643

Qy 475 VRTVEEVRETVGKLS 489

Db 644 TSGTSATAESIGNLA 658

RESULT 9

ADA33885

ID ADA33885 standard; protein; 709 AA.

XX ADA33885;

AC ADA33885;

XX DT 20-NOV-2003 (first entry)

XX Acinetobacter baumannii protein #1046.

XX Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KW plant biocontrol agent.
XX Acinetobacter baumannii.
OS US6562958-B1.
FN 13-MAY-2003.
PD 04-JUN-1999; 99US-00328352.
PF 09-JUN-1998; 98US-0088701P.
PR (GENO-) GENOME THERAPEUTICS CORP.
XX Breton G, Bush D;
XX WPI; 2003-576092/54.
DR N-PSDB; ADA29759.
XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX Example; SEQ ID NO 5172; 328pp; English.
XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX Sequence 709 AA;

Query Match 15.6%; Score 373.5; DB 6; Length 709;
Best Local Similarity 28.7%; Pred. No. 1.7e-18;
Matches 100; Conservative 67; Mismatches 141; Indels 41; Gaps 5;
QY 144 LLDLADVDVADRGEEAAAADVELVAPLPLMLLLTDFDQIAMTYIDSVARQLRHDS 203
Db 388 LLEIAD--LADGLRSTATVSE-----DFTGATADSNFADQ 424
QY 204 RQELANAVATHVEAPLSLEATSQDVAERTDTMPARTDDQVDRMADVSREISSVSASVEE 263
Db 425 LRDL-----VSRIHETSQEVARYTDTQTSITNQLAAESEHQAEIAGASTANNE 473
QY 264 VASTADV-RTSDEAEE-----LAQGEAAADALATWTDIDEATDGTGAGVEQLGER 316
Db 474 MAQSIDQVSANASAEVAQRVSQIASNGAQVNRSTEGMDTIREQIOETSKRIKLGES 533
QY 317 AADVESYTGVIDDIAETQNNLAINASTEAAPAGAGEGFAVADVKALAESEREQSTRV 376
Db 534 SQEIGNVSLINDIADQTNILANAAIQASWAGAGRGFAVADVEQVLRASASATKQI 593
QY 377 BELVEQMQAEETFTVDQLDEVQNRIGRGVVERVEAMETLQBITDAVEDAASGMQEVSTAT 436
Db 594 ETLVKTIQTDNEAVISMEQTTEVRGANLAKDAGIALDEIQKVSGLAKLIASIDAA 653
QY 437 DEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVTRVERVTV 485
Db 654 KLOSASASHIATTTVVVQETTSQTTTATFTDARSVELANNAESLRESV 702

RESULT 10
ABU40238
ID ABU40238 standard; protein; 686 AA.
XX

AC ABU40238;
XX 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #25765.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
KW Pseudomonas putida.
OS WO200277183-A2.
FN 03-OCT-2002.
PD 21-MAR-2002; 2002WO-US009107.
PF 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
PI WPI; 2003-029926/02.
DR N-PSDB; ACA44108.
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 68162; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 686 AA;

Query Match 15.5%; Score 371.5; DB 6; Length 686;
Best Local Similarity 24.7%; Pred. No. 2.3e-18;
Matches 119; Conservative 91; Mismatches 196; Indels 75; Gaps 11;

[illegible]

```

Db      19 GKXDESSNVTQAIINLYLDVLDGKEPQPLG-----LSKXDEB-- 59
Qy      162 AAVDELVARFPLMLKLTFFDQOIAMDTVIDSYAQRHDEIDSRQELANAVATHVEAPLSS 221
Db      60 -----VLKKVAFRLKRGQKKIN-----VKDKIENLKE-----VIENLEEKIG- 97
Qy      222 LEATSDQVAERTDMRATDQVDRMADVSREISVSASVEEVASTADDVVRTSEDABAL 281
Db      98 -EVLGDLGE-VNELSVRLNNENKIAEVDYIQTLSAGIEMNVQAQOLSDFALSESASM 155
Qy      282 AQCGEAAADALATMTDIDEATDGTAGVQELGERAADVESVTGVIDDIAEQTNMLALNA 341
Db      156 AEKGRQISDNVALKVSRISETSREMSDAVRLAEVSKKINDIVYVISSIASQTNLLALNA 215
Qy      342 SIEARAGEGEGFAVADEKALABESRQSTRVEELVEQQAETETVDQLDEVNORI 401
Db      216 SIEARAGEAGRGFAVAENVELADRSKSGAEIRNIIEMQ-----ENINRVIQAI 268
Qy      402 GEGVERVEEAMETLOEITDAVED-----AASGMQEVSTATDEQAVSTEVEAEVMDGVD 454
Db      269 QENVRVTEVEKAEQNLTAADFDIRRANETANMVKELSEGIDEQANSVQMLVDKIDSIS 328
Qy      455 DRAGEIAALDDIADATDQQVRTVEVRETGVKL 488
Db      329 KDVSNNLNFATQLTDTISSLEKLEKNEVEITKL 362

```

RESULT 14

ABU23101
ID ABU23101 standard; protein; 539 AA.

AC ABU23101;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #8628.

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Bordetella pertussis.

PN WO200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

PA (ELIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR N-PSDB; ACA26971.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

XX isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 51025; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC the gene product or that has an activity against a biological pathway

CC the gene product or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *X. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at

XX ftp.wipo.int/pub/published_pct_sequences

XX Sequence 539 AA;

Query Match 14.9%; Score 356.5; DB 6; Length 539;
Best Local Similarity 26.6%; Pred. No. 2.1e-17;
Matches 126; Conservative 81; Mismatches 153; Indels 113; Gaps 15;

```

Qy      38 SFTGIDDDTMAAL-----AAEQPLFEATADALVTDYDHLSEYERTQDLFANSTKTVQQL 92
Db      126 SFAASLDEMMAALERNDDAAYLQKNVKAQOASAAFAARLGEFSTNLDKLSSETLAHET 185
Qy      93 KETQAEYLLGLGRGEYDTEYAAQRAIRIKIHVDLGLGPDVYLGYTRYTYTLLDALADDV 152
Db      186 RETIMLYV-----YAA-----LLLLIVGVIAGSVLYMTRAVVRFLQR-- 222
Qy      153 VADRGEEAAAADVDELVARFLEPMLKLLTFDQIQAMDYTDVYAAQRHLHDEIDSRQELANAVA 212
Db      223 -ASQOFERMAAGD-LSARI-----EGNSRNEIGALIT 252
Qy      213 T--HVEAPLS-----SLEATSQDVAERTTMEARTDDQVDVDMADYSREISSVSA 259
Db      253 ALRHMQESLTRTVAAVRREGVDEINVGSREISAGNTDLSRTEEQ-----AASLEETAA 305
Qy      260 SVEEVAST-----ADDVRTSE---DAEALAQCGEAAADDALATMTDIDEATDGVTAGVEQ 312
Db      306 SMEQLASTVKQNADNARQANQALASASDVASGGSVSEVWSTM-----DGISASSRK 358
Qy      313 LGERAADVESVTGVIDDIAEQTNMLALNASTEARAGAGEGFVAVDEVKALABESREQ 372
Db      359 ISE-----IVSVIDGIAFQTNILALNAVAEAAARAGEQGGFVAVAGEVRSRQAOSA 411
Qy      373 STRVEELVEQQAETETVDQLDEVNQIRIGEVVEEAMETLOEITDAVEDAASGMORV 432
Db      412 AKBIKVLIE-----DSVD-----KVGTSQOVERAGATMQEIVASVKRVTDIMGEI 457
Qy      433 STATDEQAVSTEVEAEVMVDGVDRAGETAAALDDIADATDQQVRTVEEVRET 485
Db      458 SAASEQSSGIEQVNRVNSQMDVETQQNAALVEEAAAAGSLQEQAQRLAEAV 510

```

RESULT 15

ABU49142

ID ABU49142 standard; protein; 626 AA.

XX AC ABU49142;

XX 19-JUN-2003 (first entry)

DT

50
—
2
K

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:23:41 ; Search time 24.8644 Seconds
(without alignments)
1015.311 Million cell updates/sec

Title: US-09-455-978B-2

Perfect score: 2394

Sequence: 1 MSNDNDTLVADVRNGIDGH.....ATDQQRVTVBVRVGVKLS 489

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*

1: /cgn2_6/ptodata/2/iaa/5A-COMB.pap.*

2: /cgn2_6/ptodata/2/iaa/5B-COMB.pap.*

3: /cgn2_6/ptodata/2/iaa/6A-COMB.pap.*

4: /cgn2_6/ptodata/2/iaa/6B-COMB.pap.*

5: /cgn2_6/ptodata/2/iaa/PTCUS-COMB.pap.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	412.5	17.2	640	4	US-09-252-991A-23252
2	401.5	16.8	545	4	US-09-252-991A-31086
3	398	16.6	663	4	US-09-252-991A-23255
4	397.5	16.6	906	4	US-09-252-991A-32715
5	389.5	16.3	696	4	US-09-252-991A-16965
6	385	16.1	710	4	US-09-252-991A-32789
7	380.5	15.9	734	4	US-09-252-991A-30703
8	380.5	15.9	857	4	US-09-252-991A-23956
9	379	15.8	701	4	US-09-252-991A-23288
10	373.5	15.6	709	4	US-09-328-352-5172
11	373	15.6	614	4	US-09-252-991A-31412
12	368.5	15.4	573	4	US-09-252-991A-31412
13	366	15.3	760	4	US-09-252-991A-18744
14	361	15.1	613	4	US-09-252-991A-31724
15	359	15.0	579	4	US-09-252-991A-25899
16	358.5	15.0	611	4	US-09-543-681A-6665
17	352.5	14.7	563	4	US-09-252-991A-20097
18	348.5	14.6	684	4	US-09-252-991A-31048
19	346.5	14.5	504	4	US-09-252-991A-28604
20	343.5	14.3	537	4	US-09-252-991A-26180
21	334.5	14.0	653	4	US-09-252-991A-20929
22	334	14.0	572	4	US-09-543-681A-18264
23	333	13.9	645	4	US-09-252-991A-16799
24	329.5	13.8	548	4	US-09-252-991A-16799
25	321.5	13.4	535	4	US-09-543-681A-4593
26	320.5	13.4	670	4	US-09-252-991A-26867
27	315	13.2	595	4	US-09-543-681A-6908

28 313 13.1 552 4 US-09-543-681A-8191 Sequence 8191, Ap
29 313 13.1 680 4 US-09-252-991A-26639 Sequence 26639, A
30 292.5 12.2 531 4 US-08-976-063E-34 Sequence 34, Appl
31 288 12.0 519 4 US-09-543-681A-6505 Sequence 6505, Ap
32 287.5 12.0 515 4 US-09-543-681A-6739 Sequence 6739, Ap
33 270.5 11.3 485 4 US-09-252-991A-24246 Sequence 24246, A
34 263 11.0 452 4 US-09-252-991A-31873 Sequence 31873, A
35 261 10.9 472 4 US-09-543-681A-6560 Sequence 6560, Ap
36 239.5 10.0 2310 4 US-09-874-923-120 Sequence 120, App
37 231.5 9.7 454 4 US-09-252-991A-28633 Sequence 28633, A
38 221.5 9.3 107 4 US-08-428-414A-3 Sequence 3, Appli
39 205 8.6 955 2 US-08-006-676B-1 Sequence 1, Appli
40 202.5 8.5 955 1 US-08-282-845-2 Sequence 2, Appli
41 202.5 8.5 955 5 PCT-US94-00324-1 Sequence 1, Appli
42 202.5 8.5 955 5 US-08-875-435B-4 Sequence 4, Appli
43 195.5 8.2 1972 4 US-08-973-462-8 Sequence 8, Appli
44 194 8.1 1786 3 US-08-875-435B-3 Sequence 3, Appli
45 191.5 8.0 1972 4 US-08-875-435B-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-252-991A-23252
; Sequence 23252, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252.991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 23252

; LENGTH: 640

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-23252

Query Match 17.2%; Score 412.5; DB 4; Length 640;

Best Local Similarity 24.4%; Pred. No. 3.3e-24;

Matches 149; Conservative 97; Mismatches 199; Indels 165; Gaps 19;

Qy 7 TLVTADVRNGIDG-----HALADRIGL-----DEAETAWRLSFTGIDDDTWAALAAEQPLF 57

Db 59 TLTAISIQSWLEGRMHVLEGLASQLALLDQPDENIARQL-----EQPVF 103

Qy 58 -----EATADALVTDFYDHL-BSYE-RTQDLFANS-----TKTVEQ 91

Db 104 SRNPASVYLGEASGTTMPDYDAMPEGYDPRTRAWYKDALADRLIVTEFFVDAGTGEQ 163

Qy 92 -----LKEAQEYLLGLGRGEYDTE-----YAAQARAGIKI--HDVLGL- 128

Db 164 ILAMSLPVRHAGQLGVAGADMKLETTAINSILKPDGAGYAFVLSAGKILLHPDLSGLV 223

Qy 129 -----GPDVILGAY-----TRYTGLLD----- 146

Db 224 LKTLAEAYPKGAPNIVEGVHEVELDGRSQFVSFTPVKGLPGVTWYVALVLDRTATYSMLS 283

Qy 147 -----ALADDDVA-----DRG---EAAANVDLVARFLPMKLL 178

Db 284 EFTSAIVATLIVGVGIMLLGLMLRVLMQPLTDMGRAMODIAGGEGDITKR-----LKVT 339

Qy 179 TFDQQLAMDTYIDSYAQRHLHDEIDSRQELANAVATHVAPLSLEATSDQVAERTDTMRA 238

Db 340 SNDEFGLTANAFNRFVERIHESI---REVAGTA-----RQHDVAQLVVNASNSMA 388

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 32715

; LENGTH: 906

; TYPE: PR1

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-32715

Query Match 16.6%; Score 397.5; DB 4; Length 906;

Best Local Similarity 24.8%; Pred. No. 7.9e-23;

Matches 129; Conservative 87; Mismatches 158; Indels 147; Gaps 13;

QY 79 QDLFANSTKTVEQLK-----ETQAEYLLGLGRGEYDEYAAQARARIGKIHVGLGLPD 131

Db 406 QDLFNRMKPISQLKIVADAVSVVDLHKYRAGVFDEERLQBELS-----452

QY 132 VYLGAYTRYTGLDALADDVVDGRGEAAAVDELVARFLPMLKL-----178

Db 453 ---GALSRIKSWADYSAD-----HRTAAKEIIEISLPTLERVKRMTLAYGEQARA 501

QY 179 -----TFDQOI-----AMDYIID---SYAQRLLHDEIDSRQ-----206

Db 502 GSLRNVAGTFRNEMYGAFDPLGTALSTLIDQLSEGAKLNEQMKRYDSMRTFFLLICA 561

QY 207 -----LANAVATHVEAPLSLEATSQDVAERTD-TMR--ARTDDOV-----244

Db 562 AALVLIIVAAAFISLUSIMRPLSDLRGVRRVQDSSNLTIRADARGDESDTARAFNMML 621

QY 245 -----DRMADVSREISSVSAS-----VEEVASTADDDVR 272

Db 622 ESQALLRHLAETARKLITISDEMSAISNOVSHVATSGQDQDDMVATAVHQMVAQVDA 681

QY 273 RTSDEAALAOQGEAADDALATMTDIDEATDGT-----AGVEQLGERAADVESVTG 325

Db 682 RNAQAARASASANSSEHTGTGLVHANLDAIQLGLSVVMVGEAGAVDTLRNKTEESTVLE 741

QY 326 VIDIAEQTNMLNASTAEARAGEGFAVADDEVKALAEESREQSTRVEELVEQMOQA 385

Db 742 VIQIAQTNLLNLALEARAGEAGGFVADDEVKALAEESREQSTRVEELVEQMOQA 801

QY 386 ETEETVQDLDEVNORIGEGVERVEEAMETLQETITDAVEDAASGMQEVSTATDEQAVSTEE 445

Db 802 GASSAVSVMQSREQAQSQVQRAHEAGKALGLIAQAVEGIAQSNQAQISTATEEQATASE 861

QY 446 VAEMVDGVDDRAGEIA-----AALDDIADATDOQVR 476

Db 862 VSQNIIDSLNAGIEVAGAEKVTSTSSVELAKLNGLEQQIQ 902

RESULT 5

US-09-252-991A-16965

Sequence 16965, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 16965

LENGTH: 696

TYPE: PR1

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-16965

Query Match 16.3%; Score 389.5; DB 4; Length 696;

Best Local Similarity 25.3%; Pred. No. 2.3e-22;

Matches 131; Conservative 98; Mismatches 210; Indels 79; Gaps 14;

QY 1 MSNDNDTLVTADVRNGIDGHALADRI--GLDEAEIAWLRSFTGIDDDTMAALAAEQPLFE 58

Db 213 LAGDENSVOAAD-SFGRDA-SIFGRVLKMGQEGNAAMSISKV-TNAEAVDLRNETAELEFE 269

QY 59 ATADALVTDFDHLLESYERTODLF-----ANSTKTVEQLKETOQBYLL---GLGRGEYD 109

Db 270 -----FVSGSVD--EILETSPDLFQVREAAANNIFSVSTLLDKASQLADGFENLAGGRSI 322

QY 110 TEYAAQRARIGIKIHVDVLGLGPDVVLGAVTRYTG-----LIDALAD 150

Db 323 NLFA-----GVVLGALALASIIILGLVMVRETNRLLAETAKKNDNRNQAAILRLLEIAD 376

QY 151 DVADRGEEAAAVDELVARFLPMLKLLTFDQCIAMDTYIDSYAQRLLHDEIDSRQELA-- 208

Db 377 --LADGDLTVAATVE-----DFTGAIADSIYNSIDQLRELVEVET 413

QY 209 -NAVATHVEAPLSLEATSQDVAERTDTRMARTDDQVDRMADVSREISSVSASVEEVAST 267

Db 414 INQTAQVAAAAQETQSTAMEHLAE-----ASEHQAEIAGASAAINEMAVSIDQVSAN 466

QY 268 ADDVRRTSEDAEALAOQGEAADDALATMTDIDEATDGTAGVQLGERAADVESVTGVI 327

Db 467 ASESVAERVAIANKNEVHVHTITGMDNIRFQIQTSTKRIKLGESSQEIIGDIVSLI 526

QY 328 DDIAEQTNMLNASTAEARAGEGFAVADDEVKALAEESREQSTRVEELVEQMOQAE 387

Db 527 NDIADQTNILNAAAIQASMAGDAGRGFAVADDEVKALAEESREQSTRVEELVEQMOQAE 586

QY 388 EETVQDLDEVNORIGEGVERVEEAMETLQETITDAVEDAASGMQEVSTATDEQAVSTEE 447

Db 587 NEAVISMEQTTSEVVRGARLAQAQAGVALEETEKVSKTLAALIQNISNAARQAQASAGHIS 646

QY 448 EMVDGVDDRAGEIAALDDIADATDQVTVVEEVRET 485

Db 647 NTMVIQIITSQTSAGTTATARSIGNLAKMASEMRNSV 684

RESULT 6

US-09-252-991A-32789

Sequence 32789, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 32789

LENGTH: 710

TYPE: PR1

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-32789

Query Match

Best Local Similarity 27.8%; Pred. No. 5.4e-22;

Matches 131; Conservative 83; Mismatches 215; Indels 42; Gaps 10;

QY 39 FTGIDDDTMAALAAEQPLFEAT-----ADALVTDFYDHLSEYERTQDLFANS---85

Db 257 PERSKABQAFAPDALRQATTLRGQLPCEADAALQAMSGSQGREGGIEQFRAGVIR 316
Qy 86 TKTVQLKETQAEYLLGLGR-----GEYDTEYAAQARIGK-IHVDVLGLGPDVYLG-AYT 138
Db 317 TRQAQWQSSTQDMARAGRTLTEAGRQLRESTASDRASLWLIATLALAFAGCAGVAIN 376
Qy 139 RYVTGLLD-ALADDVADRGEAAADELVARFLPMLKLTFFDQOIAMDTYIDSYAQL 197
Db 377 RQIVRPLDEALA-----QAEATAAGDLGKRPONPLTLORDEL-----GQLQRV 420
Qy 198 HDET-DSRQELANAVATHVEAPLSLEATSDVAERTDTMEARTDQVDRMADYSREISS 256
Db 421 NQMGDSRLRELVRIGDGV-----SOLASSAELSAVTEQTRAGVNSQKVETDQVATAMHE 476
Qy 257 VSASVEEVASTADDVRRTSDEBALAQOGEAAADALATMTDIDBATGVTAGVEQLGER 316
Db 477 MAATVQDVARNAEASQAARQADEARQGDVQAVTRIERLASSEMDVVSSEAMARKNE 536
Qy 317 AADVSTVGVDDTARQTNMLNALSIEAARAGEGFAVVADVKALAESREQSTRV 376
Db 537 SEQIGSVLDVTKSVAEQTNLALNAIEAARAGDAGRGFAVVADVEVRGLAQRTQOSTAEI 596
Qy 377 BELVEQMAETEETVDQLDEVNORIGEGVERVEEAMETLQETDVEDAASGMQEVSTAT 436
Db 597 EGLIQLRQOQAGEAERLENSRLTASTVELARRAGALDSITRTVSDIQNNMLQIATAA 656
Qy 437 DEQAVSTVEVAMVDGVDDRAGEIAAALDDIADATDQOVRTVVEEVRETVGK 487
Db 657 EQQSTVABEINRSLVSRDVAEQSAASEQTAASSGELARLGTQLQAQVGR 707

RESULT 7
US-09-252-991A-30703
; Sequence 30703, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30703
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30703

Query Match 15.9%; Score 380.5; DB 4; Length 734;
Best Local Similarity 24.7%; Pred. No. 1.3e-21;
Matches 119; Conservative 91; Mismatches 178; Indels 93; Gaps 13;

Qy 41 GIDDDTMAALAAEQPLFEATADALVDFYDHLSEY-----ERTQDLFANSTKTV 90
Db 310 GNEVDNMLKNTVDQPLYDIDAE-----HGHIELFPLFTIADSGVRWTLMLQIPQAAVRG 363
Qy 91 QIKETQAEY-----LLGLGRGEYDTEYAAQARIGKTHDVLGLGPDVYLGAY-----T 138
Db 364 ELQQLQGLSDORQQDILGM-----SLAGLVAAALGL-LVVNLVGYGIARPL 409
Qy 139 RYVTGLLDALAD---DWDADRGEAAAAVDEL--VARFLPMLKLTFFDQOIAMDTYIDSY 193
Db 410 RQIVGMLDDIAQOEGDLTRLSSERA---DELGSIK-----GFTNFIKGL 452
Qy 194 AQLRHDIDSRQELANAVATHVEAPLSLEATSDVAERTDTMEARTDQVDRMADYSRE 253
Db 453 QNMIGVQVQSVQKVS-----DSSEHTADIAIRTNQGVQOQL---AE 490

Qy 254 ISSVSASVEEVASTADDVRR-----TSEDAEALAQOGEAAADALATMTDIDBATGV 306
Db 491 IELVATVHMTATQDVARNATHAAEAANHADQAAGKQIVSSSSRAIQALASEIGRA 550
Qy 307 TAGVEQLGERAADVESVTGVIDDIAEQTNMLNALSIEAARAGEGFAVVADVKALA 366
Db 551 VGVVONLAKSENINAILVAIRGIAEQTNLALNAIEAARAGEGFAVVADVEVRNLA 610
Qy 367 BESREQSTRVELVEQMAETEETVDQLDEVNORIGEGVERVEEAMETLQETDVEDAA 426
Db 611 QKTOQATEEIQSMITQQLQGGTRDVVKVQMSQERTDSDVRHARQAAEALESITQAVSVIN 670
Qy 427 SGMEVSTATDQAVSTVEEVAEMVDGVDDRAGEIAAALDDIADATDQOVRTVVEEVRETVG 486
Db 671 DMNFIASAAEQSAVEDINRNVANICQVANOVAGGADEASQASAEITRLAEQORRLVN 730
Qy 487 K 487
Db 731 Q 731

RESULT 8
US-09-252-991A-23956
; Sequence 23956, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23956
; LENGTH: 857
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23956

Query Match 15.9%; Score 380.5; DB 4; Length 857;
Best Local Similarity 26.8%; Pred. No. 1.6e-21;
Matches 141; Conservative 92; Mismatches 226; Indels 67; Gaps 16;

Qy 2 SNNDTLVTADVNRNGIDGHALADRIGLD-----EAEIAW--RLSFTGIDD-----DTM 47
Db 352 ANSLDILQSLAEDG--AVALKESQGHGDEPLLLQQAQVAVNRLVLQGLDEARSRL 409
Qy 48 AALAAEQP---LFEA-----TADALVDFYDHLSEYERTQDLFANSTKTVBOLK 93
Db 410 AADAQEGPKSLREALELAARLEQAITDDAVYVVVKDVLNTNIRGFADKLAE--YRASQ 467
Qy 94 ETQAEYLLGLGRGE---YDTEYAAQARIGKTHDVLGLGPDVYLGAYTRYTGLLDALA 149
Db 468 EQMTAAMGERAGQVAMRVDRSWEAQOQAM--LHS-LRTNSLLIIVGA-----AVLALLV 518
Qy 150 DDVADRGEEAAAAVDELVARFLPMLKLTFFDQOIA---MDTYIDSYAQLRHDIDSRQ 206
Db 519 -----GLGAAGFISLLIVR--PLRQAMGVARIAGBDLAVRVD-----ERRDEVGQLMA 566
Qy 207 LANAVATHVEAPLSLE-----ATSDVAERTDTMEARTDQVDRMADYSREISSYSA 259
Db 567 AMRAMTGLSRGIVSQDQGVGRIAGSEALSQVTRTRTRIGIDSQRAETEQTAVAMNQMA 626
Qy 260 SVEEVASTADDVRRTSDEBALAQOGEAAADALATMTDIDBATGVTAGVEQLGERAAD 319
Db 627 TVHEVAHNABEAGAESAESADGKVSQGVQEVVRRQTLERIERLAEVAARATASVEALSAD 686
Qy 320 VESVTGVIDDIAEQTNMLNALSIEAARAGEGFAVVADVKALAESREQSTVEEL 379

687 IGSVLDVKSVAEQTNLLNALNAEAAAGAGQGGFVAVDEVALARRTQOSTAEITL 746
380 VEQMAETEETVDQLDVNRQIGGVERVEAMETLQETIDAVEADAAGMGEVSTATDEQ 439
747 IGALQNGTQAVQRMORSHQVLDQSDVDALQTEAALGNIAFAVALIQMNQOIAAASEQ 806
440 AVSTEEVAEMVDGVDVDRAGEAALDDIADATDQQVTVVEVRET 485
807 SAVAEINRSVTAREVADQQAQWQSTASSEQALGRELQGMV 852

RESULT 9

US-09-252-991A-23288
; Sequence 23288, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23288
; LENGTH: 701
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23288

Query Match 15.8%; Score 379; DB 4; Length 701;
Best Local Similarity 29.0%; Pred. No. 1.6e-21;
Matches 96; Conservative 67; Mismatches 146; Indels 22; Gaps 4;

158 EEAATAADELVARFLPMLKLLTFQQTAMDTYDSYAQRHDEIDSQELANAVATHVEA 217
384 QDIAQGGEDLTKR---LAVTSRDEFGLGDAFNQFVERHRSI---REVAGTA----- 430
218 PLSLEATSQDVAERTDMARTDQVDRMADVSREISSVASVEEVASTADDVRRTSED 277
431 --HKLHDVSQLVNASNSMANSDEQSNRTSVAAAINELGAAQETARNADASHRASD 488
278 AELAQQEAAADALATMTDIDATDGVTVAGVQLGERAADVSVTVGVIDDIAEQTNML 337
489 ANHQAEQKQVVEQTIAMNLSSEKISASCANIEALNSRTVNIQIILEVIKIGSEQTNLL 548
338 ALNASEAARAGEGFAVVADEVKALAEESRQSTREVELVQMQMAETEETVDQDDEV 397
549 ALNAAIEAARAGEAGRFVAVDEVRLNLAHQESAQIQIKMIEELQVGAREAVATMTES 608
398 NQICEGVERVEEAMETIQETIDAVEADAAGMGEVSTATDQAVSTEEVAEMVDGVDRA 457
609 QRYSLSEVIEANRAGESLSVTRRIGEDIGNVQSVATATEQTA-----VDSLNDI 661
458 GEIAAALDDIADATDQQVTVVEVRETGKL 488
662 TEINTLNGEVNQLATIRACCELETQAGRL 692

RESULT 10

US-09-328-352-5172
; Sequence 5172, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5172
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5172

Query Match 15.6%; Score 373.5; DB 4; Length 709;
Best Local Similarity 28.7%; Pred. No. 4.3e-21;
Matches 100; Conservative 67; Mismatches 141; Indels 41; Gaps 5;

144 LLDALADDVADRGCEAAAVDELVARFLPMLKLLTFDQQTAMDTYDSYAQRHDEIDS 203
388 LLDEIAD--LADGDLRSYATVSE-----DFTGAIADSNFALDQ 424
204 RQELANAVATHVEAPLSLEATSQDVAERTDMARTDQVDRMADVSREISSVASVEE 263
425 LRDL-----VSRIHETSQEVARYTQDTQSIINQLAEASEHQAEIAGASTAMNE 473
264 VASTADDV-RTSDEAEE-----LAQQEAAADALATMTDIDATDGVTVAGVQLGER 316
474 MAQSIDQVSNASASEAERVSQVIAASGAQVNRISIEGMDTIREQIETSKRIKLGES 533
317 AADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGFVAVDEVKALAEESREQSTRV 376
534 SOEIGNIVSLINDIADQTNILALNAIQASWAGAGRGFVAVDEVORLAERSASATKQI 593
377 BELVEQMAETEETVDQLDVNRQIGGVERVEEAMETLQETIDAVEADAAGMGEVSTAT 436
594 ETLVKTIQTDNEAVISMEOQTTEVVRGANLAKDAGIALDEIQKVSGLAKLIASISDA 653
437 DEQAVSTEEVAEMVDGVDVDRAGEAALDDIADATDQQVTVVEVRET 485
654 KLOSASASHIATMTVVQEIETSTTTATFTARSVSELANMAESLRESV 702

RESULT 11

US-09-252-991A-31412
; Sequence 31412, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31412
; LENGTH: 614
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31412

Query Match 15.6%; Score 373; DB 4; Length 614;
Best Local Similarity 26.6%; Pred. No. 3.8e-21;
Matches 142; Conservative 87; Mismatches 214; Indels 90; Gaps 18;

9 VTADVNRGIDGHALADRIQ-----LDEAEIAWRLSFTG-IDDDTMAALAAEQPLFEA 59
108 LTGDVR-----AVRELLGGPVRAAQQLIDEA---NLQFRGQVQEWKNVLLRGQRQ--- 152
60 TADALVTDFVDHLESYER-TQDIFANSTKTVQKQAEYLLGLGRGEY---DTEVAAQ 115
153 -AEA-QTKYQSFEAQAEQAVQDILGRIGSVAGELKDRVERL-----REEHRLGTAYRQG 206
116 RARIGKTHDVLGLGPDVYLG-----AVTRYVTGLLDALADDV-----VADR- 156

Db 207 RQRE-----LEAGADPIAGQAVGTGIDRATTAAQMALRDELHQASDLHSSISAEARRT 260
QY 157 -----GEEAAAANDE--LVARFL--PMLKLTTPQOQIAMDYIISYARLHDEIDS 203
Db 261 MLLGSLVLIGASLAVALLSLMLVNRNLVRPQVRLIEHIAQLSHGDFGERIEIRKDELCK 320
QY 204 RQELANAVATHEAPLSLEATSQDVAERTDT-----MRARTDQVDRMADVSRREISS 256
Db 321 LALAANTURDLVDIFDLRRSTRDLSDASGSLNIAISLMAAGTREQFSRTDQVATAMQE 380
QY 257 VSASVEEVASTADVRRRTSEDAEALAAQGEAAADALATMTDIDEATDGTAGVQLGER 316
Db 381 MSATAQEVARYAGDAARAADDEADDSQORGDMVEETIRSIGEMRKEIDHTVEVIRQLSD 440
QY 317 AADVESVTGVDDIDIAEQTNMLNALSIEAARAGEGFAVVADRVKALAEESRQSTRV 376
Db 441 SGRIGKVLVDIVGTAEQTNLLALNAAIEAARAGDAGRGFAVVADRVTLAORTAESIAEI 500
QY 377 EELVEQMAETEETVDQDEVNQRIGEGVERVEEAMETLOEITDAVEDAASGMQEVSTAT 436
Db 501 HQIDTVQNGAVNAARAIESGOSRSEAGABQVANAGAMLQITASVESIRDMNRQIATAA 560
QY 437 DEQAVSTEEVAMVDGVDRAGEIAAALDDIADATDQQVTVVEEVRETVGKLS 489
Db 561 BEQTAVEDISGRNLT-----EIAS-----IASSNOEQVEQTEAASRDHLGLS 602

RESULT 12
US-09-252-991A-18744
; Sequence 18744, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18744
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18744

Query Match 15.4%; Score 368.5; DB 4; Length 573;
Best Local Similarity 25.0%; Pred. No. 7.9e-21;
Matches 129; Conservative 89; Mismatches 175; Indels 123; Gaps 15;
QY 14 RNIGDGHALADRI-GL--DEABIAWRLSFTGTGDDTMAALAAEQPLFEATAD-ALVTDIFY 69
Db 100 RSLLEGHLSAEQIAGLRNDHAEINLQAL-----DRVQYAAQMP---GEAELAKVAEP- 149
QY 70 DHLSEYETQDLFANSTKTVQEKETOAEYLLGLGRGYDTEYAAQARIGIKIHVDVLGLG 129
Db 150 -----ERGYALWSATSARVLSLAASDFPSAAQLSYGSDRQFGAMREVINQL----- 196
QY 130 PDVYLGAVTRYTGLLDALDADVDRGEAAAAVDELVARFL-----PMLKLTTPDQ 183
Db 197 -----DEMEEAANAADGASGALSALGERHRWQOVALVAFGLL 231
QY 184 IAMDY-----IDSYAQRLLH-----EIDSRQE-----LANAVATHV 215
Db 232 VCLSLVLVFPGLVTRPLQRLLEIANGDGLRVLRLVLTSTRDEPGRIGLSAFNAFLDKL 291
QY 216 EAPLSLSEATSQDVAERTDTMPARTDQVDRMAD-----VSREISSVSASVEEVAST 267
Db 292 QPLIREVGRVTEVADSAGSLAGMTAAN--DRLINSEHASVDPQVSTAQTMSSAVHEVARN 350

QY 268 ADDVRTTSEDAEALAAQGEAAADALATMTDIDEATDGTAGVEQLGERAADVESVTGVI 327
Db 351 AQSAQVADDARRQAREGANVVEATIEVIRQLAQEVSSSESIQQLAQETASIDAVLTVI 410
QY 328 DDIAEQTNMLNALSIEAARAGEGFAVVADRVKALAEESRQSTRVEELVEQMAQET 387
Db 411 KGIAEQTNLLALNAAIEAARAGEQGRGFAVVADRVKALAEESRQSTRVEELVEQMAQET 470
QY 388 BETVDQLDEVNQRIGEGVERVEEAMETLOEITDAVEDAASGMQEVSTATDEQAVSTEEVA 447
Db 471 QNAVAMOSGSLKARDSVER-----AAGVDGVLAAT-----G 502
QY 448 EMVDGVDRAGEIAAALDDIADATDQQVTVVEEVRE 483
Db 503 DAVGRINDLAAQIASACEEQSRVIDEAIARNISEVRE 538

RESULT 13
US-09-252-991A-31724
; Sequence 31724, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31724
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31724

Query Match 15.3%; Score 366; DB 4; Length 760;
Best Local Similarity 31.7%; Pred. No. 1.8e-20;
Matches 97; Conservative 61; Mismatches 126; Indels 22; Gaps 4;
QY 195 QRL--HDEIDSRQELA---NAVATHVEAPLSLEATSQDVAERTDTMRTATDQVDRMADV 250
Db 454 QRLPHTGRDELGELAGWFNRFLDKLPRIIRVVKSVDRDARSTADQSAISSQTSAGMQQQ 513
QY 251 SREISSVSASVEEVASTADVRRRTSEDAEALAAQGEAAADALATMTDIDEATDGTAG- 309
Db 514 FREIDQVATASHMTATTAQDVARSAAQAADAARGADQATRDGLALIDRTTQSIDSLSAANL 573
QY 310 -----VEQLGERAADVESVTGVDDIDIAEQTNMLNALSIEAARAGEGFAVVADRVK 363
Db 574 TSAMQGVBEQLASSSEETIGSVLEIRATAEQTNLLALNAAIEAARAGDAGRGFAVVADRV 633
QY 364 ALAEESRQSTRVEELVEQMAQETEETVDQDEVNQRIGEGVERVEEAMETLOEITDAVE 423
Db 634 NLARTQSVQIRGVIEGLQGTTRDVVDAMHGHRHQAGSVQVEQVDEAVALQRICEAVT 693
QY 424 DAASGMQEVSTATDEQAVSTEEVAMVDGVDRAGEIAAALDDIADATDQQVTVVEEVRE 483
Db 694 VINDMNLQIASAAHEQSSVAEEINRNV-----AAIRDVTESLSQAEEESAQVSQ 742
QY 484 TVGKLS 489
Db 743 SLNRLA 748

RESULT 14
US-09-252-991A-25899
; Sequence 25899, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:28:46 ; Search time 71.0412 Seconds

(without alignments)
2159.184 Million cell updates/sec

Title: US-09-455-978B-2

Perfect score: 2394

Sequence: 1 MSNDNTLTADVRNGIDGH.....ATDQQRVTEEVRETVGKLS 489

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	399.5	16.7	535	12	US-10-282-122A-66393
2	399.5	16.7	535	12	US-10-389-647-531
3	386.5	16.1	682	12	US-10-282-122A-66174
4	374.5	15.6	680	12	US-10-282-122A-69670
5	371.5	15.5	686	12	US-10-282-122A-68162
6	361	15.1	644	12	US-10-282-122A-77591
7	359.5	15.0	891	9	US-09-272-809-5
8	356.5	14.9	539	12	US-10-282-122A-51025
9	348.5	14.6	679	12	US-10-389-647-372
10	348	14.5	626	12	US-10-282-122A-77066
11	338	14.1	686	12	US-10-282-122A-52305
12	332	13.9	547	12	US-10-332-288-28
13	326	13.6	641	12	US-10-282-122A-77278
14	315	13.2	643	12	US-10-282-122A-77602
15	314.5	13.1	501	12	US-10-282-122A-69564

16	313.5	13.1	541	12	US-10-282-122A-77047
17	308	12.9	845	12	US-10-282-122A-76517
18	292.5	12.2	531	8	US-08-976-063C-34
19	292.5	12.2	531	11	US-09-750-986D-34
20	292	12.2	564	12	US-10-335-977-6156
21	289	12.1	564	12	US-10-335-977-6157
22	289	12.1	565	10	US-09-882-227-462
23	284	11.9	630	10	US-09-882-227-460
24	280	11.7	675	12	US-10-335-977-4944
25	276	11.5	293	12	US-10-335-977-4942
26	273.5	11.4	664	12	US-10-335-977-4943
27	269	11.2	753	12	US-10-282-122A-47230
28	263	11.0	431	12	US-10-389-647-451
29	257	10.7	431	12	US-10-012-819-228
30	254	10.6	433	8	US-08-945-038-6
31	252.5	10.5	433	12	US-10-335-977-8500
32	252.5	10.5	438	12	US-10-335-977-8501
33	245	10.2	883	15	US-10-369-493-18563
34	241	10.1	654	12	US-10-282-122A-76558
35	241	10.1	673	12	US-10-335-977-6249
36	239.5	10.0	2310	9	US-09-874-923-120
37	239.5	10.0	2310	9	US-09-991-496-120
38	239.5	10.0	2310	10	US-09-820-843A-114
39	231.5	9.7	2354	10	US-09-820-843A-113
40	225	9.4	1190	15	US-10-369-493-18546
41	221	9.2	566	12	US-10-243-552-921
42	217.5	9.1	202	12	US-10-335-977-6246
43	212	8.9	1258	14	US-10-156-761-10395
44	212	8.9	2124	15	US-10-369-493-11841
45	209	8.7	533	12	US-10-282-122A-49144

ALIGNMENTS

RESULT 1

US-10-282-122A-66393
; Sequence 66393, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636

```
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66393
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66393

Query Match      16.7%; Score 399.5; DB 12; Length 535;
Best Local Similarity 25.2%; Pred. No. 1.8e-18;
Matches 132; Conservative 111; Mismatches 206; Indels 75; Gaps 16;

QY 3 NNDTTLVTADVNRGIDGHALADRLGDEAEIATWLSFTGIDDDTMAALAAEQPLFEATAD 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 SENELSVNA-LRNHMEGDMHD-----ALRADVLAAFWV-QPGDGAAG 78

QY 63 ALVTDFVHLESYERTQDLFANSTKTVEQ-----LKETAQAEYLLGLGRGYDTEYAAQRA 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 79 QVRQDLQHSQWFR-----KVVEQNOGLPLNDIAHOALVEL-RPDLEAVIGAAES 127

QY 118 RIGK-IHDLVLGLGPDV--YLGAYTRYTGLLDALADDV-----VADRGE----- 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 IVGKALLDPVAARAEALPQFVQAF-KELEGRNEALSSLIEKHVEQTNRAEDSMRYSAWML 186

QY 160 -----AAAADVELVARPL-----PMLKLLTFDQOIAMDTY-----IDS--YAQRLHDEI 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 AGGILVACLVLGQLCQLLRAVLQPLRLKLVSSARVIAQGNLQEPITGVDSNDEAGLQAL 246

QY 202 DSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRATDDQVDRMADVSEISSVSASV 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 GEMQENLRQMTITIRQSEELHDTSSQSIGQTSQSVHVGASQQADSATSMAASMEIMTNI 306

QY 262 EVASTADDDVRTSEDAAALAQOGEAAADALATMTDIDEATDGTAGVEQLGERAADVE 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 SQISDHADNARVISAKSEELASSGGQVILNVVEGMSRIADVNNQSSITSITAGQSSDEIH 366

QY 322 SVTGVDDIDIAQTNMLNALSIEAARAGEGFAVVADEKALAEBSRGSTVEELVE 381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 367 SIIQVIRGIAEQTNLLNALNAIEAARAGEAGRGFAVVADEVRGLAARTTQSTQBITAMIE 426

QY 382 QMQAETETVDQDEVNQRIGEGVERVEEAMETLQETITDAVEDAASGMQEVSTATDEQAV 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 427 RIRASTQAINSMAGVSRVNEGVSFAQGVSNIEILDGTRHAASVVDEISQITREQSR 486

QY 442 STEEVAEMVGVDDVRAGEIAAALDDIADATDQQVRTVEEVRETV 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 487 ASDEIAQRVELIAQRSQNTQAMHEMA-AT---ARRLNEVAATM 526

RESULT 2
US-10-389-647-531
; Sequence 531, Application US/10389647
; Publication No. US20040033549A1
; GENERAL INFORMATION:
; APPLICANT: GREENBERG, E. Peter
; APPLICANT: SCHUSTER, Martin
; APPLICANT: LOSTROH, Candi
; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
; FILE REFERENCE: UIZ-038CP
; CURRENT APPLICATION NUMBER: US/10/389,647
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 09/653730
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/153022
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 710
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 531
; LENGTH: 535
```

```
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-389-647-531

Query Match      16.7%; Score 399.5; DB 12; Length 535;
Best Local Similarity 25.2%; Pred. No. 1.8e-18;
Matches 132; Conservative 111; Mismatches 206; Indels 75; Gaps 16;

QY 3 NNDTTLVTADVNRGIDGHALADRLGDEAEIATWLSFTGIDDDTMAALAAEQPLFEATAD 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 SENELSVNA-LRNHMEGDMHD-----ALRADVLAAFWV-QPGDGAAG 78

QY 63 ALVTDFVHLESYERTQDLFANSTKTVEQ-----LKETAQAEYLLGLGRGYDTEYAAQRA 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 79 QVRQDLQHSQWFR-----KVVEQNOGLPLNDIAHOALVEL-RPDLEAVIGAAES 127

QY 118 RIGK-IHDLVLGLGPDV--YLGAYTRYTGLLDALADDV-----VADRGE----- 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 IVGKALLDPVAARAEALPQFVQAF-KELEGRNEALSSLIEKHVEQTNRAEDSMRYSAWML 186

QY 160 -----AAAADVELVARPL-----PMLKLLTFDQOIAMDTY-----IDS--YAQRLHDEI 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 AGGILVACLVLGQLCQLLRAVLQPLRLKLVSSARVIAQGNLQEPITGVDSNDEAGLQAL 246

QY 202 DSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRATDDQVDRMADVSEISSVSASV 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 GEMQENLRQMTITIRQSEELHDTSSQSIGQTSQSVHVGASQQADSATSMAASMEIMTNI 306

QY 262 EVASTADDDVRTSEDAAALAQOGEAAADALATMTDIDEATDGTAGVEQLGERAADVE 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 SQISDHADNARVISAKSEELASSGGQVILNVVEGMSRIADVNNQSSITSITAGQSSDEIH 366

QY 322 SVTGVDDIDIAQTNMLNALSIEAARAGEGFAVVADEKALAEBSRGSTVEELVE 381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 367 SIIQVIRGIAEQTNLLNALNAIEAARAGEAGRGFAVVADEVRGLAARTTQSTQBITAMIE 426

QY 382 QMQAETETVDQDEVNQRIGEGVERVEEAMETLQETITDAVEDAASGMQEVSTATDEQAV 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 427 RIRASTQAINSMAGVSRVNEGVSFAQGVSNIEILDGTRHAASVVDEISQITREQSR 486

QY 442 STEEVAEMVGVDDVRAGEIAAALDDIADATDQQVRTVEEVRETV 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 487 ASDEIAQRVELIAQRSQNTQAMHEMA-AT---ARRLNEVAATM 526

RESULT 3
US-10-282-122A-66174
; Sequence 66174, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
```



```
Db 199 SFKIEDTGFVFLTNAOGEVQIHRQKEQV-----KSLQQIYSGSALLNKGSGFNLLSTDY 254
Qy 113 AAQRARIGKIHDLVGLGPDVYLGAYTRYTGLL---DALAD-DVADR---GEEAAAAYD 165
Db 255 QGEVWVASI-----YIEMDFELVGTVPVHEVFAELDAVAGQWMLTLVAIAIF 304
Qy 166 ELVARFLPMLKLLTFDQIQIAMDYIDSYAQRLLHD-----BIDSRQELIA----- 208
Db 305 IFMIGIFLA-----NSIAMP--INQIAKRFTDLGRGDGDLRSORIEVKNGNDEIAQLSKG 354
Qy 209 -NAVATHEAPLSLEATSQD-----VAERTDTMEARTDDQVDRWADYSRETSVSSAS 260
Db 355 FNGFIEKIHOSIKDVQTSRELQVAAGVSKALVTHDMSQQQQRDTQIQVTAIINQMGAT 414
Qy 261 VEEVASTADDVRRTSSEDAALAOQGEAAADALATMTDIDBATDGTAGVEQLGERAADV 320
Db 415 ISEIASNAATAETANQASGNADQGRNVVVKAKAISLAHDIENTGKVVEQLASTTQEI 474
Qy 321 ESVTGVDDIDIAEQTNMLNALSIEAARAGEGFAVVADEVKALAEESRQSTRVEELV 380
Db 475 GSILDAIRGISEQTNLLNALNAIEAARAGDQGRGFAVVADEVRLASRTASSTEEIQKMI 534
Qy 381 EQMAQEETVDQDEVNQRIGEVVEREAMETLOEITDAVEDAASGMQEVSTATDEQA 440
Db 535 NQLQNDAKNAVSAAMDAGKTVTHQGVAAASDEAVQVLSISDRHIDLSDRNTQVATATERQS 594
Qy 441 VSTEVEAMVGDVDRAGEIAAALDDIADATDQOVRTVEEVRETVGKL 488
Db 595 TVVHTINQIEINAINEVTTSTAEELADAS-----KSLRELSGRL 635

RESULT 7
US-09-272-809-5
; Sequence 5, Application US/09272809
; Patent No. US2002002239A1
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C.
; TITLE OF INVENTION: Phytotoxins as fluorescent labels
; FILE REFERENCE: 2500.118U50
; CURRENT APPLICATION NUMBER: US/09/272, 809
; CURRENT FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 5
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: cph5 locus SLL0041
; OTHER INFORMATION: (locus 1001300) an 891 aa protein,
; OTHER INFORMATION: methyl-accepting chemotaxis protein I. Homology
; OTHER INFORMATION: to tsr in last 250 aa.
US-09-272-809-5

Query Match 15.0%; Score 359.5; DB 9; Length 891;
Best Local Similarity 27.1%; Pred. No. 1.5e-15;
Matches 136; Conservative 86; Mismatches 197; Indels 83; Gaps 19;

Qy 23 ADRIGLDEAIEAWRLSFTGIDDDTMAALAAEQP--LFEATADALVTDVFDHLESY----- 75
Db 418 ADRV-----IVRFQATWAGTVVSVSABGYKPAKGATADPCFADSY--VEKYRSGR 469
Qy 76 ERTQDLFANSTKT---VEQLK--ETQAEYLLGLGRGEYDTEYAAQARIGKIHDLVGL-- 128
Db 470 QATRDYI-NAGLTPCHIGQLKPEVXANLV-----APINYKGNLLGLLI 512
Qy 129 -----GP-----DVYLGATRYTYTGLL-----DALADDVVADGGERAAAVDELVARFLP 173
Db 513 AHQCSGRPDWHQNEIDLFGQLTVQVGLALERSDLLAQOKIAE--VEQQRQKRMQKRALE 570
Qy 174 MLKLLTFDQIQIAMDTYIDTSVAQRLLHDEIDSRQELANAVATHVEAPLSLSLEATSQDVART 233
```

```
Db 571 L--LMEVDPVSRGDLTIRAHV--TEDEIGTIADSYNATIESLRRIVTQVTAASQFTETT 626
Qy 234 DT-----MRARTDDQVDMADVSRFISVSASVEVASTADDVRRTSSEDAEALAOQGE 286
Db 627 DTNEVAVRQAQANQALDVAELERLQAMNKSIOAENAAQAESAVERATQTVDOGE 686
Qy 287 AAADDALATMTDIDEATDGTAGVEQLGERAADVESVTGVDDIDIAEQTNMLNALSIEAA 346
Db 687 DAMRTVDGIVATRETVAAATAKQVKRLGESSQKSVNVLIGSFADQTNLLNALNAIEAA 746
Qy 347 RAGEAGEGFAVVADEVKALAEESRQSTRVEELVEQMAETEETVDQDEVNQRIGEGVE 406
Db 747 HAGEGREGFAVVADEVRSRLARQSAEATAEIAQLVATTIQAETNEVYNVNAMEAGTEQVVVGTK 806
Qy 407 RVBEAMETLOEITDAVEDAASGMQE--VSTATDEQAVSTEE-----VAEMVD----- 451
Db 807 LVETRESLQNQIT-AVSAQISGLVEALTSAAIEQSQISESVTQTMALVIAQIADKNSSEAS 865
Qy 452 GVDDRAGEIAAALDDIADATDQ 473
Db 866 GVSATFKELLAVASQLQEAQVQ 887

RESULT 8
US-10-282-122A-51025
; Sequence 51025, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 51025
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Bordetella pertussis
US-10-282-122A-51025
```

✓

US-10-282-122A-77066

Query Match 14.5%; Score 348; DB 12; Length 626;
Best Local Similarity 27.4%; Pred. No. 5.7e-15;
Matches 92; Conservative 78; Mismatches 134; Indels 32; Gaps 6;

QY 174 MLKLLTPDQI-----AMTYIISYAQRLHDEIDSRQELANAVATHVEAPLSS 221
Db 298 MSPLKTLDSAIKDIAASGGDLTKLIDNLDKFEFSELALGFNSFTMLGSIQIRQLKTIASG 357
QY 222 LEATSQDVARTDTRMARTDDQVDRMADVSRREISSVSASVEEVASTADDVVRTSEDAEAL 281
Db 358 VLGAETANEAEVSRILVEQOL-----QELQOLATANMEMANTASEVANSQVRAADA 410
QY 282 AQCGEAAADALATMTDIDEATDGT-----AGVEQLGE--RAAD-VESVTGVDDIA 331
Db 411 AKEGESA--SLEGGSSVHHTTDAIQLRSIRIGSSVEDVKELVATRIETVLDVINDIA 467
QY 332 EQTNMLAINASIEARAGEGFAVVADEVKALAEBSREOSTRVEELVEQMQAETEETV 391
Db 468 DQTNLLALNAIEAARAGESGRGFAVVADEVRTTLAQRTOQSTMQISEIEQLQEGAKNVS 527
QY 392 DQDEVNQIGEGVERVEEAMETIQEITDAVEDAASGMQEVSTATDEQAVSTEVEAMVD 451
Db 528 RSMDESKLETVIVKTNQVNEKISLVQQAHLIRSDMNLQIASAABEQSILVAEELNNTV 587
QY 452 GVDDRAGEIAAALDDIADATDQVQRTVEEVRETGVK 487
Db 588 NIKLSIKLSEASAGNAGTEMNAQSVKVEQNELLNE 623

RESULT 11

US-10-282-122A-52305

Sequence 52305, Application US/10282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 52305
LENGTH: 686
TYPE: PRT
ORGANISM: Clostridium botulinum
US-10-282-122A-52305

Query Match 14.1%; Score 338; DB 12; Length 686;
Best Local Similarity 26.8%; Pred. No. 3e-14;
Matches 91; Conservative 80; Mismatches 130; Indels 38; Gaps 4;

QY 178 LTFDQIQIAMDYIDSYAQLRHDEIDSRQELANAV-----ATHVEAPLSSLE---ATSQ 227
Db 350 LTKVEIWNDEIDGKLSKIFNTMIDSLREITRNINFSIQLAGSSQELSSAEQTSVASE 409
QY 228 DVAERTDTRMARTDDQVDRMADVSRREISSVSASVEEVASTADDVVRTSEDAEALAQGEA 287
Db 410 EISSATEEIASGAENQVKASNESSLLMNDVMGNMYTLKEEFDEIISFNNTNTLASKQE 469
QY 288 AADALATMTDIDEATDQVGTAGVQLGERAADVESVTGVDDIAEQTNMLAINASIEAAR 347
Db 470 NMSNMVQOMATIKNSVNVSSNIMYDLOKNSSEIGNIVEIINTIADQTNLAINASIERAR 529
QY 348 AGEAGEGFAVVADEVKALAEBSREQSTRVEELVEQMQAETEETVDQDLDENVNRIGEGVER 407
Db 530 AGEAGKFAVVADEVKRLAEESINSANNIKNLINWTDQKTALNSIKDGASQSEKGESI 589
QY 408 VEEAMETIQEI-----TDAVEDAASGMQEVSTATDEQAVSTEVEV 446
Db 590 VAEVKESLGEILNGFSNVNHKFAVSVDMSITASNDISITAMASKLYDIETISNTASANTEEV 649
QY 447 AEMVDGVDDRAGETAAALDDIADATDQVQRTVEEVRETGV 485
Db 650 AASTE-----EOSATIEEITESIEKLVSMVENLKESV 681

RESULT 12

US-10-332-288-28

Sequence 28, Application US/10332288
Publication No. US20040054165A1

GENERAL INFORMATION:

APPLICANT: RAINNEY, Paul Barton
APPLICANT: SPIERS, Andrew Julien
APPLICANT: BANTINAKI, Eleni
APPLICANT: BANTINAKI, Eleni

TITLE OF INVENTION: BACTERIAL POLYSACCHARIDE AND BIOFILM DEVELOPMENT

FILE REFERENCE: I0317.70012US00

CURRENT APPLICATION NUMBER: US/10/332,288

CURRENT FILING DATE: 2003-10-06

PRIOR APPLICATION NUMBER: PCT/GB01/03077

PRIOR FILING DATE: 2001-07-09

PRIOR APPLICATION NUMBER: UK 0016842.7

PRIOR FILING DATE: 2000-07-07

NUMBER OF SEQ ID NOS: 34

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 28

LENGTH: 547

TYPE: PRT

ORGANISM: Pseudomonas fluorescens

US-10-332-288-28

Query Match 13.9%; Score 332; DB 12; Length 547;
Best Local Similarity 25.0%; Pred. No. 5.6e-14;
Matches 119; Conservative 84; Mismatches 189; Indels 84; Gaps 12;

QY 66 TDFYDHLESYERTQDLFANSTKVEQLKETQAEYLLGLRGGEYDTYAA-----114
Db 89 TDKNDYKSFARIEQQMANVEXTI-----HGQAD-----RMEFDNFKAHINYKNVLAQV 138
QY 115 -QPARIGKTHDVGLGPDVYLVGAYTRYTYTGLLDALADDV-VADRGEEAAAADVDELV--- 168
Db 139 LERVEANDLPGANQLLEEQLTPIWTEGRMKLNDIITENKNVSDR---ATAADEAVLSAK 195

QY 169 -----ARFLPMLKLLTFDQOIAMDTVID--SYAORLH-----DE 200
 Db 196 ISMAVSLIIAIALAGLGLLMLRAIMAPMQRIVDI-----LETWRDGLSKRLNLERKDE 250
 QY 201 IDSRQELANAVATHVERPLSSLEATSQDVARTDTMRAATDDQVDRMADVSRISSVSAS 260
 Db 251 FGAVETGFNDMMTELIALVSQAORSSVQVTTSTVEIAATSKQQAATATATAITTEIGAT 310
 QY 261 VERVASTADDVVRTSDEAALAQGEAAA---DIALA---TMTDIDEATDGTAGVEQL 313
 Db 311 SREIAATSKDLVMTVEVSTAAQAQSVAGSGGQGLARMEETWHSVWGAAADLVNAKLAIL 370
 QY 314 GERAAVESVTGVIDDIAETQNNMLALNASIEAARAGEAGEGFVAVADEVKALAEESREOS 373
 Db 371 NEKAGNINQVVTVIKVADQTNLLSLNAIEAEKAGEYGRGFVAVATEVRLADQATVAT 430
 QY 374 TRVEELVEQMAETETVPOLDENVORIGEVVERSEAMETLOEITDAVEDAASGMQEV 433
 Db 431 YDIEQVREIQSAVSAGVGMGDKFSEVVRGMEFVQVQVGEQSLQIIHQVALAPRVLVNV 490
 QY 434 TATDEQAVSTEEVAEMVVDGVDRAGEIAAALDDIADATDQVRTVEEVRETGVKLS 489
 Db 491 EGMQAQATGAQINH-----ALVOLGASSQTVESLRQASFAIDELS 532

RESULT 13

US-10-282-122A-77278
 ; Sequence 77278, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR FILING DATE: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 77278
 ; LENGTH: 641
 ; TYPE: FRT
 ; ORGANISM: Vibrio cholerae

US-10-282-122A-77278

Query Match 13.6%; Score 326; DB 12; Length 641;
 Best Local Similarity 24.1%; Pred. No. 1-7e-13;
 Matches 117; Conservative 97; Mismatches 164; Indels 108; Gaps 16;
 QY 45 DTMAALAAEOPLEAFATADALVTFDYPHLRSYERTQDLFANSTKTVQLKETQAE----- 98
 Db 203 DTVRTLLAEAKNAANIESLIKDISNGAE-----MVSNGVKSSQKAESYQKGHVNV 255
 QY 99 YLILGCRGVDTEYAAQRAIRIGKHIDVILGIPDPVILGAYTRYVTGLLDAL-ADVVADRG 157
 Db 256 KQTLIRNE-----MVDLVKD-----SEVLLKADEMSSAAVIALKGSSEVSQA 300
 QY 158 EAAAAVDELVARFLPMLKLLTFDQOIAMDTVIDSYAQRHLHIDSRQELANAVATHVEA 217
 Db 301 QEQSAACEE-----SLKSLD-QQOIALDGAUTA-AQSLDELITDEL- 339
 QY 218 PLSSLEATSQDVARTDTMRAATDDQVDRMADVSRISSVSASVEVASTADVVRTSE- 276
 Db 340 -----TSTDIVKSSSEVAAAEEELSAGIEENKSSNEIMGALNQISSGACHMAKSVET 392
 QY 277 -----DAALAQOGEAAADAL-----ATMTDIDE-----ATDGVTAGVEQLGE 315
 Db 393 GITSLTQIEQARLGKERAESLKACEKMLTGIEENKTTVDEMILAITSTKAATENLNE 452
 QY 316 RAADVSTGTVDDIAB-----QTNMLALNASIEAARAGEAGEGFVAVADEVKALAE 368
 Db 453 -MANIERISRQDKIVDGISNVSIQTAMLVANGAVEAARAGEYKGFVAVSTDIONLAND 511
 QY 369 SREQSTRVEELVEQMAE-----TTE-----TVOLDENVORIGEGVER 407
 Db 512 AAENAEQIKQVKNIQEQINIVRKDLADILSTWMBEAAQKALTIKQLDNVSRMSDVILG 571
 QY 408 VEEAMETLQEIITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVVDGVDRAGEIAALDDI 467
 Db 572 SKRISESAGGIERSTADARAGMQOITATAESSHATGEAATAARQSSSTSELASAIENI 631
 QY 468 ADATDQ 473
 Db 632 AAVADE 637

RESULT 14

US-10-282-122A-77602
 ; Sequence 77602, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR FILING DATE: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77602
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-10-282-122A-77602

Query Match 13.2%; Score 315; DB 12; Length 643;
Best Local Similarity 24.9%; Pred. No. 9.2e-13;
Matches 100; Conservative 77; Mismatches 167; Indels 58; Gaps 10;

QY 133 YLG-AYTRYTGLDADLADD-----VVADGEEAAAADDELVAR----- 170
Db 249 YLGEFDGYKRV--NLTDMMYTLMVFDKRA--TALADVDVAVANAITTGIVLIGISVGV 305
QY 171 -----FLPMLKLLFDQIQAMDTYIDSVAQRLHDEIDSRQELAN----- 209
Db 306 IIFIINQIKPFLKRLKVAVLDAQS--GDLTRRL--EVNGNDLQAISEGFNRFSGNLQ 361
QY 210 ----AVATHVEAPLSSLEATSQDVAERTDTMTARTDDQVDMADVRSISVSASVEVA 265
Db 362 NMMLQISDATQIISSEIQQLSQ-TAKENEQMLISHSETDQ---VVTALTQMSSESAR-TVA 417
QY 266 STADDVRTSDEAALQGEAAADDALATMTDIDEATDGTAGVEQLGERAAADVSTVG 325
Db 418 ESVTQSNQITAAKSEAQQSLVIVNAVSTVTSLVNVDENMSSESINMNRDANKISEVLS 477
QY 326 VIDDIAETQNTMLNASTAARAGAGGFAVADVKALAESEQSTRVEELVEQMOA 385
Db 478 VIGALSEQTNLALNAAIEAARAGQGRGFAVADVKALARTQNSTEISDMLTKLE 537
QY 386 ETEETVDQLDEVNQRIQEGVERVEEAMETLQBITDAVEDAASGMQEVSTATDEQAVSTEE 445
Db 538 GTDSVVGAMERTKQCQTADKTSEVSGSLNMSASVSDDLDLSTQIAAATEQOSTVAAE 597
QY 446 VAEWVDGDDDRAGEIAALDDIADATDQOVRTVEEVRETGVK 487
Db 598 LSRNMLSIRESLVSQRTVGATESLSHNSHLEQLVGK 639

RESULT 15
US-10-282-122A-69564
; Sequence 69564, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69564
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas syringae
US-10-282-122A-69564

Query Match 13.1%; Score 314.5; DB 12; Length 501;
Best Local Similarity 27.4%; Pred. No. 7.3e-13;
Matches 86; Conservative 59; Mismatches 136; Indels 33; Gaps 5;

QY 176 KLLTFDQIQAMDTYID-----SYAQRHDEIDSRQELANAVATHVEAPLSSLEATSQD 228
Db 184 RMLATTNQFSPDSRIDERSGKHVSLAQRFEQFLAQITGLVDGVVRDTR----GLGELGHD 239
QY 229 VAERTDTMR-----ARTDDQVDMADVRSREISS-VSASVEEVASTADDVVRTSED 277
Db 240 LAKASGTLETGAQHQLSEIARMTGAMQRMGMAMNDISGHVAVQVORAGDASDV----- 293
QY 278 AEALAQOGEAAADDALATMTDIDEATDGTAGVEQLGERAAADVSTVGVIDDIAEQTNML 337
Db 294 -----AHGRDSVDRAQSEITQLAARISTTDETQVQALANQSEIQIGKVLVDVIGSIAEQTNLL 348
QY 338 ALNASTAARAGAGGFAVADVKALAESEQSTRVEELVEQMOAETEETVDQLDEV 397
Db 349 ALNASTAARAGAGGFAVADVKALAESEQSTRVEELVEQMOAETEETVDQLDEV 408
QY 398 NQRIQEGVERVEEAMETLQBITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRA 457
Db 409 LQGVGRCVENSQRASELSRVSVEGEGIGCHITQLNGLIATTEQQTATASREIADQLRSVQAIA 468
QY 458 GEIAAALDDIADAT 471
Db 469 EHTAANIGVLAASS 482

Search completed: August 10, 2004, 15:42:26
Job time : 73.0412 secs

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:22:45 ; Search time 21.3123 Seconds
(without alignments)
2207.061 Million cell updates/sec

Title: US-09-455-978B-2

Perfect score: 2394

Sequence: 1 MSNDNDTLVTADVVRNGIDGH.....ATDQQQRTVEVRETVGKLS 489

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2394	100.0	489	2 T44978	transducer protein
2	2360	98.6	489	2 B84304	Htr10 transducer [
3	643	26.9	481	2 A84294	Htr9 transducer [i
4	609.5	25.5	482	2 T44973	transducer protein
5	597	24.9	633	2 H84336	Htr3 transducer [i
6	597	24.9	805	2 T48840	transducer protein
7	580	24.2	790	2 B84238	Htr18 transducer [
8	578	24.1	778	2 F84237	Htr4 transducer [i
9	573	23.9	778	2 T48897	transducer protein
10	560.5	23.4	452	2 T44849	Htr5 transducer [i
11	555	23.2	810	2 F84327	halobacterial tran
12	554	23.1	810	2 T46810	transducer protein
13	540.5	22.6	777	2 T44597	sensory rhodopsin
14	537.5	22.5	534	2 S55299	Htr15 transducer [
15	535.5	22.4	636	2 A84252	transducer protein
16	529	22.1	804	2 T44606	transducer protein
17	524	21.9	451	2 T44964	transducer protein
18	523.5	21.9	544	2 T44938	transducer protein
19	523.5	21.9	643	2 H84305	Htr8 transducer [i
20	521.5	21.8	628	2 F84219	Htr16 transducer [
21	521	21.8	545	2 E84327	Htr7 transducer [i
22	521	21.8	545	2 T46811	halobacterial tran
23	509.5	21.3	789	2 E84236	Htr6 transducer [i
24	507.5	21.2	788	2 T44262	transducer protein
25	506	21.1	420	2 C84298	Htr12 transducer [
26	505	21.1	642	1 T44253	transducer protein
27	502.5	21.0	773	2 T44989	transducer protein
28	499.5	20.9	627	2 F84194	Htr14 transducer [
29	497.5	20.8	419	2 T44276	transducer protein

ALIGNMENTS

RESULT 1

T44978

transducer protein hemAT [validated] - Halobacterium salinarum

N:Alternate names: methyl-accepting taxis protein htrB; transducer protein htrB; transduce:

C:Species: Halobacterium salinarum

C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 15-Sep-2000

C:Accession: T44978

R:Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.

Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996

A:Title: Signal transduction in the archaeon Halobacterium salinarum is processed throu

A:Reference number: Z22804; MUID:96209786; PMID:8643458

A:Accession: T44978

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-489 <ZHA>

A:Cross-references: EMBL:U75436; NID:gl654420; PIDN:AAB17881.1; PID:gl654421

A:Experimental source: strain Flx15

A>Note: the source is designated as Halobacterium salinarum

C:Genetics:

A:Gene: hemAT; htp15; htrB

C:Function:

A:Description: involved in aerotactic signal transduction; involved in oxygen sensing; "

C:Superfamily: Halobacterium salinarum transducer protein htr1

C:Keywords: heme; methylated amino acid; signal transduction

Query Match 100.0%; Score 2394; DB 2; Length 489;
Best Local Similarity 100.0%; Pred. No. 7.8e-98;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSNDNDTLVTADVVRNGIDGHADRLGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT	60
DB	1	MSNDNDTLVTADVVRNGIDGHADRLGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT	60
QY	61	ADALVTDFYDHLSEYERTQDLFANSTKTVEQLKETAQYLLGLGRGEYDTEYAQRARIG	120
DB	61	ADALVTDFYDHLSEYERTQDLFANSTKTVEQLKETAQYLLGLGRGEYDTEYAQRARIG	120
QY	121	KIHDLVGLGPDVYLGATRYTITGLLDALADDVVADRGEAAAVDELVARFLPMLKLTTF	180
DB	121	KIHDLVGLGPDVYLGATRYTITGLLDALADDVVADRGEAAAVDELVARFLPMLKLTTF	180
QY	181	DDQIAMDTYIDTSVAQRLLHDEISRQELANAVATHVEAPLSLSLEATSQDVAERTDTMEART	240
DB	181	DDQIAMDTYIDTSVAQRLLHDEISRQELANAVATHVEAPLSLSLEATSQDVAERTDTMEART	240
QY	241	DDQVDRMADYSREISSVSASVEEVASTADVRRTSEDAEALAAQGEAAADALATMTDID	300
DB	241	DDQVDRMADYSREISSVSASVEEVASTADVRRTSEDAEALAAQGEAAADALATMTDID	300
QY	301	EATDGVGTAGVEQLGERAADVESVTGVDIAEQTNMLALNASIEAARAGEGFAVVAD	360
DB	301	EATDGVGTAGVEQLGERAADVESVTGVDIAEQTNMLALNASIEAARAGEGFAVVAD	360

QY 361 EVKALAEESRQSTRVEELVQMQAETETVDQLDEVNQRIEGVERVEEAMETIQEITD 420
 Db |||||
 QY 361 EVKALAEESRQSTRVEELVQMQAETETVDQLDEVNQRIEGVERVEEAMETIQEITD 420
 QY 421 AVEAASGMQEVSTATDEQAVSTVEAVMVDGVDDRAGEIAAALDDIADATDQQVTVVEE 480
 Db |||||
 QY 421 AVEAASGMQEVSTATDEQAVSTVEAVMVDGVDDRAGEIAAALDDIADATDQQVTVVEE 480
 QY 481 VRETGVKLS 489
 Db |||||

RESULT 2

E84304
 Htr10 transducer [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: E84304
 R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
 A;Title: Genome sequence of Halobacterium species NRC-1.
 A;Reference number: A84160; MUID:20504483; PMID:11016950
 A;Accession: E84304
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-489 <STO>
 A;Cross-references: GB:AE004437; NID:g10580997; PIDN:AAG19801.1; GSPDB:GN00138
 C:Genetics:
 A;Gene: htr10
 C:Superfamily: Halobacterium salinarum transducer protein htr1

Query Match 98.6%; Score 2360; DB 2; Length 489;
 Best Local Similarity 99.0%; Pred. No. 2.4e-96;
 Matches 484; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MSNDNDTLTVADVRNGIDGHADRLIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
 Db |||||
 QY 1 MSNDNDTLTVADVRNGIDGHADRLIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
 QY 61 ADALVTDFDHLSEVTERODLFANSTKTVEQLKETAQAEYLLGLGRGEYDTEYAAQARIG 120
 Db |||||
 QY 61 ADALVTDFDHLSEVTERODLFANSTKTVEQLKETAQAEYLLGLGRGEYDTEYAAQARIG 120
 QY 121 KIHDLVGLGPDVYLGAITYRYTGLDALADDVVDGRGEAAAVDELVARFLPMLKLTFF 180
 Db |||||
 QY 121 KIHDLVGLGPDVYLGAITYRYTGLDALADDVVDGRGEAAAVDELVARFLPMLKLTFF 180
 QY 181 DQIQIAMDYIYSYQRLHDEIDSRQELANAVATHVEAPLSSLEATSDQVARETDTMRART 240
 Db |||||
 QY 181 DQIQIAMDYIYSYQRLHDEIDSRQELANAVATHVEAPLSSLEATSDQVARETDTMRART 240
 QY 241 DDQVDRMADVRSREISSVSASVEEVASTADDVVRTSDEAALAAQOGEAAADALATMTDID 300
 Db |||||
 QY 241 DDQVDRMADVRSREISSVSASVEEVASTADDVVRTSDEAALAAQOGEAAADALATMTDID 300
 QY 301 EATDGVTAGVQLGERAADVESVTGVIDDIAEQTNMLNALSIEAARAGEGEGFAVVAD 360
 Db |||||
 QY 301 EATDGVTAGVQLGERAADVESVTGVIDDIAEQTNMLNALSIEAARAGEGEGFAVVAD 360
 QY 361 EVKALAEESRQSTRVEELVQMQAETETVDQLDEVNQRIEGVERVEEAMETIQEITD 420
 Db |||||
 QY 361 EVKALAEESRQSTRVEELVQMQAETETVDQLDEVNQRIEGVERVEEAMETIQEITD 420
 QY 421 AVEAASGMQEVSTATDEQAVSTVEAVMVDGVDDRAGEIAAALDDIADATDQQVTVVEE 480
 Db |||||
 QY 421 AVEAASGMQEVSTATDEQAVSTVEAVMVDGVDDRAGEIAAALDDIADATDQQVTVVEE 480
 QY 481 VRETGVKLS 489

Db |||||
 481 VRETGVKLS 489

RESULT 3

A84294
 Htr9 transducer [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 17-May-2002
 C:Accession: A84294
 R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
 A;Title: Genome sequence of Halobacterium species NRC-1.
 A;Reference number: A84160; MUID:20504483; PMID:11016950
 A;Accession: A84294
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-481 <STO>
 A;Cross-references: GB:AE004437; NID:g10580901; PIDN:AAG19717.1; GSPDB:GN00138
 C:Genetics:
 A;Gene: htr9
 C:Superfamily: Halobacterium salinarum transducer protein car
 Query Match 26.9%; Score 643; DB 2; Length 481;
 Best Local Similarity 37.6%; Pred. No. 2.1e-21;
 Matches 182; Conservative 72; Mismatches 196; Indels 34; Gaps 10;
 QY 15 NGIDGHADRLIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADAL-VTFDYLHLE 73
 Db |||||
 QY 15 NGIDGHADRLIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADAL-VTFDYLHLE 73
 Db 19 NTLVDVPFAF--VDADGAVVAV-----DDQIAAL-----LETAPEDAIGVTDIGERLN 63
 QY 74 SYERTODLFANSTKTVEQLKETAQAEYLLGLGRGEYDTEYAAQARIGKTHDLVGLGPDVY 133
 Db |||||
 QY 74 SYERTODLFANSTKTVEQLKETAQAEYLLGLGRGEYDTEYAAQARIGKTHDLVGLGPDVY 133
 Db 64 D-----DGSFALANKVADTPTIDAHHEY---DGVGLADESYALLTGVDYVEDTTVAGNTDLW 116
 QY 134 LGAYTRYTGLDALADDVVDGRGEAAAVDELVARFLPMLKLL-----TFDQOI--- 184
 Db |||||
 QY 134 LGAYTRYTGLDALADDVVDGRGEAAAVDELVARFLPMLKLL-----TFDQOI--- 184
 Db 117 FIATPVYHTGEPFGVI-EIVQDR-SSSARYQSELQALFGLVDTLTDAYDAGRFDAVDIA 174
 QY 185 AMDTVIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSDQVARETDTMRARTDQV 244
 Db |||||
 QY 175 AEDTLDDDEYIQIGRNLTETFGDTLAAHITTEVNDVERLEAASQAVSESSAEIDELSTAQS 234
 QY 245 DRMDVRSREISSVSASVEEVASTADDVVRTSDEAALAAQOGEAAADALATMTDIDEATD 304
 Db |||||
 QY 235 TNVSTVATEVETLSATVQELASTADEVVDTSATERLADDGSAASDADMMADVATAAD 294
 QY 305 GVTAGVQLGERAADVESVTGVIDDIAEQTNMLNALSIEAARAGEGEGFAVVADDEVKA 364
 Db |||||
 QY 295 SVTSDEALQNRIDEDWDVITGIAEQTNMLNALSIEAARAGEGEGFAVVADDEVKA 354
 QY 365 LAEESRQSTRVEELVQMQAETETVDQLDEVNQRIEGVERVEEAMETIQEITDAVED 424
 Db |||||
 QY 355 LAEASQWAGHIESLVSEIQRTADTVDLTDTTTRIEDAVAQVEDAMASFEETIATVEA 414
 QY 425 AASGMQEVSTATDEQAVSTVEAVMVDGVDDRAGEIAAALDDIADATDQQVTVVEEVEET 484
 Db |||||
 QY 415 TAEGIEQVSDATNQASAEIEAAWDETADLADDDITTVAVDIVSQTEAASMLHLDDES 474
 QY 485 VGKL 488
 Db 475 VSEL 478

RESULT 4

T44973
 transducer protein hcpIII [similarity] - Halobacterium salinarum
 N;Alternate names: methyl-accepting taxis protein hta; transducer protein hta; transducer
 C:Species: Halobacterium salinarum
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Mar-2001

C;Accession: T44973
R;Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.
Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996
A;Title: Signal transduction in the archaeon Halobacterium salinarum is processed through a transducer protein car
A;Reference number: Z22804; PMID:96209786; PMID:8643458
A;Accession: T44973
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-482 <ZNA>
A;Cross-references: EMBL:U75435; NID:gl654418; PIDN:AAB17880.1; PID:gl654419
A;Experimental source: strain Flx15
A;Note: The source is designated as Halobacterium salinarum
C;Genetics:
A;Gene: htpIII; htp3; hta
C;Superfamily: Halobacterium salinarum transducer protein car
C;Keywords: methylated amino acid; signal transduction
F;219-472/Region: MCP signalling domain similarity

Query Match 25.5%; Score 609.5; DB 2; Length 482;
Best Local Similarity 37.2%; Pred. No. 6.2e-20;
Matches 181; Conservative 70; Mismatches 197; Indels 39; Gaps 12;

Qy 15 NGIDGHALADRLGDEAEIAWLSTGTGDDDTMAALAAEQPLFEATADAL-VTDFVDHLE 73
Db :
19 NTLDVPAPFA--VDADGVAVW-----DDQIAL-----LETAPEDAIGVTDIGERLN 63
Qy 74 SYERTQDLFANSTKTVEQLKETQAELVLGLGRGEYDTEYAAQRARICKIHDLVGLGPDVY 133
Db :
64 D----DGSRALKANKVADTPIDAHHEY---DGVLADESYALLTGDIYVEDTTVAGNTDLW 116
Qy 134 LGAYTRYVTGLDALADDVVADRGEBAARVDELVARFLPMLKL-----TFDQOI--- 184
Db 117 FIATPYHTGFEGVI-EIVQDR-SSSARYQSELQALFGELVDTLDAYDAGRFDATVDIA 174
Qy 185 AMDTYIDSVAQRHLDEIDRSQELANAVATHVEAPLSLEATSQDAERTDTMEARTDDOV 244
Db :
175 AEDTLDDDEYIQIRNLTEFTGDITLAHIITEVHVNDVERLEAASQAVSESAEIDELSTAQS 234
Qy 245 DRMADVSRISSVASVEEVASTADVRRTSDEAELAQQEAADALATMTDIDEATD 304
Db :
235 TNVSTVATEVELSATVQEIASADEVDVTSATAERLADDSRAASADAAMWADVATAAD 294
Qy 305 GVTAGVEQLGERAADVESVTGVIDDIAEQTNMLNALNASIEAAPAGEGFAVVADEVKA 364
Db 295 SVTSDVEALQNRIEDIEVDVDTGIAEQTNMLNALNASIEAAPGEEGEGFAVVADEVKA 354
Qy 365 LAESREQSTRVELVEQMCAETEETVDQLDEVNQRTGEGVERVEEAMETLQEIITDAVE- 423
Db :
355 LAEDAQSNAGHIESLSEIQRDTADTVLTDTTRIEDAVAQVEDAWASPEEIVTAVEA 414
Qy 424 --DAASGMQEVSTATDEQAVSTEEVABMVGDVDDRAGEIAAALDDIADATDQQVRTVEEV 481
Db 415 TABRASNMM--VSDATNEQAASAHEIAAMVDETADLADDDITTAVADIYSQTAEASAMLHD 472
Qy 482 RETVGKLL 488
Db 473 DESVSEL 479

RESULT 5
H84336
Htr3 transducer [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 17-May-2002
C;Accession: H84336
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablonka, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: H84160; PMID:20504483; PMID:11016950
A;Accession: H84336

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-633 <STO>
A;Cross-references: GB:AE004437; NID:gl0581300; PIDN:AAG20060.1; GSPDB:GN00138
C;Genetics:
A;Gene: htr3
C;Superfamily: Halobacterium salinarum transducer protein htrII

Query Match 24.9%; Score 597; DB 2; Length 633;
Best Local Similarity 33.8%; Pred. No. 3e-19;
Matches 150; Conservative 76; Mismatches 158; Indels 60; Gaps 7;

Qy 42 IDDDTMAALAEQPLFEATADALVTFDFDHLESYERT----QDLFANSTKTVQELKETQA 97
Db 213 LDELDPCAFGASLSQMHTRLEALITDDEAREDAEQTRKDAEEAASERLNERLERRAA 272
Qy 98 EYLLGLGRGEYDTEYAAQARIRKITHVLGLGPDVYLGYATRYTYTGLLDALADDVVADRG 157
Db 273 -----EYSDewAAAA-----GDLTR-----RLDEDVDSEPM 239
Qy 158 EEA AAAAVDELVARFLFPLMLKLLTFFDQIAMDITVIDSYAQLRHOEIDSQELANAVATHVEA 217
Db 300 QDIAEAFNDMMG-----DVEATL-----AQVRSIADAVD-----AASDTVST 336
Qy 218 PLSLEATSQDVAERTDTMRATDDQVRMADVSRSEISSVSASVEEVAATDADVRTSED 277
Db 337 SAAEIRASDQVSESVQDISADAOQORRLGTGVGDEVTSLSATVEEIASADDAVSTVQ 396
Qy 278 AEALAQCGEAAAADALATMTDIDEATDGTAGVEQLGERAADVESVTGVIDDIAEQTNML 337
Db 397 AATESRQGEIGEDAVAELERIEATADS AVERVTAL EEA VDAIGDVTGVTITDIAEQTNML 456
Qy 338 ALNASTAEARAGEGEGFAVVADEVKALAEBSREGSTRVEELVEQMAETEETVDQLDEV 397
Db 457 ALNANIEARADKSGDGFVAVDEVKDLADEKESATEIETLLVDDVQADVADTVADMSEL 516
Qy 398 NORIGEGVERVEEAMETLQEIYDAVEDASGQCEVSTATDEQAVSTEEVAEVDGVDDRA 457
Db 517 GDRVDAGSETIEAALAAALDDIGQVEAANGSVQISDATDEQAASVEEVMTIDEVTDLS 576
Qy 458 GEIAAALDDIADATQOQVTVEEV 481
Db 577 DRTATESQOVSAAEQQAASVSEV 600

RESULT 6
T48840
transducer protein bast [validated] - Halobacterium salinarum
N;Alternate names: chemotaxis transducer protein bast; methyl-accepting taxis protein ht
C;Species: Halobacterium salinarum
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Sep-2000
R;Accession: T48940; T44981
R;Kokoeva, M.V.; Oesterhelt, D.
Mol. Microbiol. 35, 647-656, 2000
A;Title: Bast, a membrane-bound transducer protein for amino acid detection in Halobacte
A;Reference number: Z24542; MUID:20138366; PMID:10672186
A;Accession: T48940
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-805 <KOK>
A;Cross-references: EMBL:AJ245950; PIDN:CAB82572.1
R;Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.
Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996
A;Title: Signal transduction in the archaeon Halobacterium salinarum is processed throu
A;Reference number: Z22804; MUID:96209786; PMID:8643458
A;Accession: T44981
A;Status: preliminary;
A;Molecule type: DNA
A;Residues: 'M',13-128,'R',130-133,'S',135-200,'MP',203-241,'R',243-332,'MR',335-401,403
<ZHA>
A;Cross-references: EMBL:U75437; NID:gl654422; PIDN:AA817882.1; PID:gl654423
A;Experimental source: strain Flx15 derivative of S9


```

QY 232 RTDTMARTDDQVRMADVRSISVSASVEEVASTADDVRRTSSEDAEALAAQGEAAAD 291
Db 498 AVGRIIRAADQDLEAVASSETDEMSATIEEVAASADQVAETSQRAAALGDDGQAAQD 557
QY 292 ALATMTDIDEATDGVTVAGVQLGERAADVESVTGVIDDIAEQTNMLNALSIEAARAGEA 351
Db 558 AVAQLEIEDETOAAATAVDDLEANKMSEIETIIVAAITDIAEQTNMLNALSIEAARADQD 617
QY 352 GEGFAVVADEVKALAEBSREQSTRVEELVEQMAETETVDQLDEVNQRIQEGVERVEEA 411
Db 618 GDGFAVVADEVKLADESASAEIEALVAEVRQATSTVAAMDRIQERVSQGVETVSET 677
QY 412 METLOETDAVEDAASQMGVSTATDEQAVSTVEVAEMVQDQVDRAGEIAAALDDIADAT 471
Db 678 ERSLSIAGRIAEADTGVEISNAMDDQAASVSDVTTAVGDVAALGEETATEAESTADAA 737
QY 472 DQQVTVVEEV 481
Db 738 AEQATTLSDV 747

RESULT 9
transducer protein VI (htp6, htd) [imported] - Halobacterium salinarum
C:Species: Halobacterium salinarum
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Sep-2000
C:Accession: T48897
R:Rudolph, J.; Nordmann, B.; Storch, K.F.; Gruenberg, H.; Rodewald, K.; Oesterhelt, D.
FEMS Microbiol. Lett. 139, 161-168, 1996
A:Title: A family of halobacterial transducer proteins.
A:Reference number: Z24094; MUID:96275896; PMID:8674984
A:Accession: T48897
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-778 <RUD>
A:Cross-references: EMBL:X95590; NID:g1435133; PIDN:CAA64843.1; PID:g1435134
A:Experimental source: strain S6
C:Genetics:
A:Gene: htpVI
C:Superfamily: Halobacterium salinarum transducer protein htrII

Query Match 23.9%; Score 573; DB 2; Length 778;
Best Local Similarity 33.0%; Pred. No. 4.2e-18;
Matches 144; Conservative 74; Mismatches 168; Indels 50; Gaps 6;

QY 88 TVEQLKETOAYLLGLGRGYDTEYAAQRA-RIGKIHDLGLGPDVYLGAAYTRYTYGLLD 146
Db 320 TVRALNDLEAK-AAALERGEYDTLDVARVDELGRLEAFASLRDTVQAIRDANEQQVD 378
QY 147 -----ALADDVVADRGEEAAAVDELVARFL-----PMLKLLTFDQOIAMD 187
Db 379 AEARGEAEAAQADAEAAQAEAEAREESEQAARLETTAEAFSETMRAYAAGDLTVRLD 438
QY 188 TYDTSVAQ-----RLHDEIDSRQELANAVATHVEAPLSLEAT 225
Db 439 ADVEQAAMADIAAFNEMAADMEATIADAGRFADAVAT-----ASTDASGAAAVEQT 491
QY 226 SQDVAERTDMRTDQVDRMADVRSISVSASVEEVASTADDVRRTSSEDAEALAAQGG 285
Db 492 GRDVSQAVGRIIRAADQRIQLEAVASETDEMATIEEVAASADOVAETSQRAAALGDDG 551
QY 286 EAAADALATMTDIDEATDGVTVAGVQLGERAADVESVTGVIDDIAEQTNMLNALSIEA 345
Db 552 QAAQDAVAQLEIEDETOAAATAVDDLEAKMSEIETIIVAAITDIAEQTNMLNALSIEA 611
QY 346 ARAGEAGEGFAVVADEVKALAEBSREQSTRVEELVEQMAETETVDQLDEVNQRIQEGV 405
Db 612 ARADQDGGFAVVADEVKLADESASAEIEALVAEVRQATSTVAAMDRIQERVSQGV 671
QY 406 ERVEEAMETLOEITDAVEDAASQMGVSTATDEQAVSTVEVAEMVQDQVDRAGEIAAALD 465
Db 672 ETVSETERSEIAGRIAEADTGVEISNAMDDQAASVSDVTTAVGDVAALGEETATEAE 731
```

```

QY 466 DIADATDQQVTVVEEV 481
Db 732 STADAAAEQASTLSDV 747

RESULT 10
T44849
transducer protein car [validated] - Halobacterium salinarum (strain S9)
N:Alternate names: arginine transducer protein Car; transducer protein htp11; transducer
C:Species: Halobacterium salinarum
A:Variety: strain S9
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Sep-2000
C:Accession: T44849
R:Storch, K.F.; Rudolph, J.; Oesterhelt, D.
EMBO J. 18, 1146-1158, 1999
A:Title: Car: a cytoplasmic sensor responsible for arginine chemotaxis in the archaeon Ha
A:Reference number: Z22859; MUID:99164082; PMID:10064582
A:Accession: T44849
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-452 <STO>
A:Cross-references: EMBL:AJ132321; NID:g4469245; PIDN:CAB38318.1; PID:g4469246
A:Experimental source: strain S9
C:Genetics:
A:Gene: car; htp11; htrXI
C:Function:
A:Description: involved in chemotactic signal transduction; arginine sensor; mediates th
C:Superfamily: Halobacterium salinarum transducer protein car
C:Keywords: methylated amino acid; signal transduction

Query Match 23.4%; Score 560.5; DB 2; Length 452;
Best Local Similarity 31.3%; Pred. No. 7.9e-18;
Matches 159; Conservative 88; Mismatches 170; Indels 91; Gaps 13;

QY 10 TADVRNGIDGHALADRI-----GLDPAEIA-----WRLSFTGIDDDTMAALAA 52
Db 5 SSDMGGEATGEHLADELCEAYLGDNEDDGGDLQRLSRERDFWKMFN-----QLVA 56
QY 53 EOP--LFEATADALVTD-----FYDHLESY-----ERTODLFANSTKTVEQLKETOAYL 100
Db 57 EYPEGILITAAADGTVTWNERFSDHMKWARSALGEDASDVFS-----TAE-SETLPEAV 111
QY 101 LGLGRGYDTEYAAQRA-RIGKIHDLGLGPDVYLGAAYTRYTYGLLDALADDVVADRGEEA 160
Db 112 VTGDTVEEE-----PHDVPTDSLCOYHGVLRAPTG-----DVVGSFG--V 152
QY 161 AAADVDELVARFLPMLKLLTFDQOIAMDVTYDSVQRLHDEIDSRQELANAVATHVEAPLS 220
Db 153 VPDISEKVKXN-----QRELHD-----LHETVSSNVGHELS 182
QY 221 SLEATSQDVAERTDMRTDQVDRMADVRSISVSASVEEVASTADDVRRTSSEDAE 280
Db 183 EISESIDVGSFAETEFAGEIEERNEGVADVSNQSAETIEEIASAEVSOASQRAQD 242
QY 281 LAQGEAAADALATMTDIDEATDGVTVAGVQLGERAADVESVTGVIDDIAEQTNMLN 340
Db 243 RATEGQTAEATIDRMGAVQESAERVNDTIDGLTSQADEMSEIIDAINDIADQTNMLN 302
QY 341 ASTEAAARAGEGFAVVADEVKALAEBSREQSTRVEELVEQMAETETVDQLDEVNQRI 400
Db 303 ASTEAAARAGEGFAVVADEVKLAESQERADEIEQIMVEMVETTDQTDADRIGQTTT 362
QY 401 IGGVVERVEAMETLOEITDAVEDAASQMGVSTATDEQAVSTVEVAEMVQDQVDRAGEI 460
Db 363 IEEAITAVRETLDSLOEIRKAVDETATGVKEVAGATDDHAASTEQVNAATTDVAVDKLT 422
QY 461 AAALDDIADATDQQVTVVEEVTVGKL 488
Db 423 EDRLNLSQIASQEHQDRVABIEDMVDEL 450

RESULT 11
```

F84327
 Htr5 transducer [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: F84327
 R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon, J.; K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A;Title: Htr5, a family of halobacterial transducer proteins
 A;Reference number: 84160; PMID:20504483; PMID:11016950
 A;Accession: F84327
 A;Status: preliminary
 A;Molecule type: DNA
 A;Cross-references: GB:AE004437; NID:g10581215; PIDN:AAG19986.1; GSPDB:GN00138
 C:Genetics:
 C:Superfamily: Halobacterium salinarum transducer protein htrII

Query Match 23.2%; Score 555; DB 2; Length 810;
 Best Local Similarity 32.1%; Pred. No. 2.7e-17;
 Matches 142; Conservative 85; Mismatches 174; Indels 42; Gaps 8;

QY 84 NSTKTVQLKETOAEYLLGLGRGEYDTEYAAQR-----ARIGKIHVDVLGLGPDVYL 135
 DB 346 NTAAGVQSLSAARAE-----IEAGNYVDVASSRRDEIGQLFASIGMRDALVTQIDAEA 401
 QY 136 AVTRYTYTGLLDALAD-----DVADRGEEAAAADDELVA-----RFLPMLKLL----- 178
 DB 402 AREQATEAQDAERERAEADAKADAEALAELEAQAERYSDVMAACADGDLT 461
 QY 179 -----TFDQOIA--MDTVDSYAQLHDEIDSRQELANAVAT---HVEAPLSSLEATS 226
 DB 462 RMPADDTTNEAMAAIAASFNEMLAQWEHTIIDI-GEFADAVATASEAEVGAADAEAS 520
 QY 227 QVARTDTMRTDQVDRMADVRSISVSASVEEVASTADDVVRTSEDAEALAOQGE 286
 DB 521 GQVSESVQETAGAADQRNMLDTVSGEMTDLSAAIEEVAASADSVAEHSHQTAETARDGE 580
 QY 287 AAADDALATMTDIDEATGVTAGVQLGERADVSVTGVIDDIABQTNMLANASIEAA 346
 DB 581 QTAEDAIERSLVQEAIDATQNVNEALDDQMAEISEIVDLISDIAEQTNMLANANIEAA 640
 QY 347 RAGEAGEGFAVVADEVKALAEBSREQSTVEELVQMQAETETVDQLDENVNQRIGEGVE 406
 DB 641 RADKSGDGFVAVVADEVKOLAEETQESAGDIERRITEVQSTTATVAEARAABESMDAGID 700
 QY 407 RVEEMETLQETDAVEDAASQGVSTATDQAVSTEEVAVEMVDGVDVDRAGEIAAALDD 466
 DB 701 AVEEVYDAFTAVSDHADETDTCVQELISDTTDDQAASTEAEVAVSMTEEVADLSDSTAGEAQS 760
 QY 467 IADATDQVTVVEEVRETGVKLS 489
 DB 761 VSAAREEQASMSSEISDSVESLS 783

RESULT 12
 T46810
 halobacterial transducer protein IV [imported] - Halobacterium salinarum
 C:Species: Halobacterium salinarum
 C>Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 02-Sep-2000
 C:Accession: T46810
 R;Rudolph, J.; Nordmann, B.; Storch, K.F.; Gruenberg, H.; Rodewald, K.; Oesterhelt, D.
 FEBS Microbiol. Lett. 139, 161-168, 1996
 A;Title: A family of halobacterial transducer proteins.
 A;Reference number: 224094; PMID:96275896; PMID:8674984
 A;Accession: T46810
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-810 <RUD>
 A;Cross-references: EMBL:X95589; NID:g1435130; PIDN:CAA64841.1; PID:g1435131

A;Experimental source: strain S9
 C:Genetics:
 A;Gene: htrIV
 C:Superfamily: Halobacterium salinarum transducer protein htrII

Query Match 23.1%; Score 554; DB 2; Length 810;
 Best Local Similarity 32.1%; Pred. No. 3e-17;
 Matches 142; Conservative 85; Mismatches 174; Indels 42; Gaps 8;

QY 84 NSTKTVQLKETOAEYLLGLGRGEYDTEYAAQR-----ARIGKIHVDVLGLGPDVYL 135
 DB 346 NTAAGVQSLSAARAE-----IEAGNYVDVASSRRDEIGQLFASIGMRDALVTQIDAEA 401
 QY 136 AVTRYTYTGLLDALAD-----DVADRGEEAAAADDELVA-----RFLPMLKLL----- 178
 DB 402 AREQATEAQDAERERAEADAKADAEALAELEAQAERYSDVMAACADGDLT 461
 QY 179 -----TFDQOIA--MDTVDSYAQLHDEIDSRQELANAVAT---HVEAPLSSLEATS 226
 DB 462 RMPADDTTNEAMAAIAASFNEMLAQWEHTIIDI-GEFADAVATASEAEVGAADAEAS 520
 QY 227 QVARTDTMRTDQVDRMADVRSISVSASVEEVASTADDVVRTSEDAEALAOQGE 286
 DB 521 GQVSESVQETAGAADQRNMLDTVSGEMTDLSAAIEEVAASADSVAEHSHQTAETARDGE 580
 QY 287 AAADDALATMTDIDEATGVTAGVQLGERADVSVTGVIDDIABQTNMLANASIEAA 346
 DB 581 QTAEDAIERSLVQEAIDATQNVNEALDDQMAEISEIVDLISDIAEQTNMLANANIEAA 640
 QY 347 RAGEAGEGFAVVADEVKALAEBSREQSTVEELVQMQAETETVDQLDENVNQRIGEGVE 406
 DB 641 RADKSGDGFVAVVADEVKOLAEETQESAGDIERRITEVQSTTATVAEARAABESMDAGID 700
 QY 407 RVEEMETLQETDAVEDAASQGVSTATDQAVSTEEVAVEMVDGVDVDRAGEIAAALDD 466
 DB 701 AVEEVYDAFTAVSDHADETDTCVQELISDTTDDQAASTEAEVAVSMTEEVADLSDSTAGEAQS 760
 QY 467 IADATDQVTVVEEVRETGVKLS 489
 DB 761 VSAAREEQASMSSEISDSVESLS 783

RESULT 13
 T44597
 transducer protein htrVI [similarity] - Halobacterium salinarum
 N;Alternate names: methyl-accepting transducer protein htrD; transducer protein htrp
 C:Species: Halobacterium salinarum
 C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Sep-2000
 C:Accession: T44597
 R;Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.
 Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996
 A;Title: Signal transduction in the archaeon Halobacterium salinarum is processed through a family of transducer proteins.
 A;Reference number: 222804; PMID:96209786; PMID:8643458
 A;Accession: T44597
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-777 <ZHA>
 A;Cross-references: EMBL:U75438; NID:g1654424; PIDN:AAB17883.1; PID:g1654425
 A;Experimental source: strain mFLX15
 A;Note: the source is designated as Halobacterium salinarum
 C:Genetics:
 A;Gene: htrpVI; htrp6; htrD
 C:Superfamily: Halobacterium salinarum transducer protein htrII
 C;Keywords: methylated amino acid; signal transduction; transmembrane protein
 F;494-746/Region: MCP signalling domain similarity

Query Match 22.6%; Score 540.5; DB 2; Length 777;
 Best Local Similarity 32.4%; Pred. No. 1.1e-16;
 Matches 141; Conservative 74; Mismatches 171; Indels 49; Gaps 6;

QY 88 TVEQLKETOAEYLLGLGRGEYDTEYAAQRA-RIGKIHVDVLGLGPDVYLGYATRYTGLD 146
 DB 320 TVRALNDLEAK-APALERGEYDLDVAVRDELGRLEAFASLRDAVQRAQSDANEQGV 378

QY 147 ALA-----DDVADRGEERAAAAVDELVARFLP 173
Db 379 AEAARSEARAAQPEAAQAEAAEESDAGPPGDDRRGLGDACVAAGDLTVR--- 435
QY 174 MLKLLTFDQQA-----MDTYIDSVQAQRHLDEIDSRQELANAVA---THVEAPLSLEATS 226
Db 436 ----LDADVEQAAMADMRFAENFNAADMEATIAERGAFADEVATASTDASAAAVEQTG 491
QY 227 ODVAERTDTMRARTDDQVDRMADYSREISSVSASVEEASTADDVRRRTSEDAEALAOQGE 286
Db 492 RDVSTAVGRIKRRADQDQLENAVSETDEMSATIEEVAVAGQVAETSQRAAALGDDGQ 551
QY 287 AAADDALATMTDIDBATDGTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAA 346
Db 552 AAAQDAVAQLEIEDETQAAATAVDDLEAKNGSEIETIVAAITDIAEQTNMLALNANIEAA 611
QY 347 RAGEAGEFVAVADVVKALAEBSREQSTRVELVQMQAETEETVDQLDEVNQRIGEGVE 406
Db 612 RADQDGDGFVAVADVVKALADESKALAEPMQALVAEVRQAQTSTVAAMDRIQERVSDGVE 671
QY 407 RVEEAMETLOETTDADVEAAQMGQEVSTATDREQAVSTVEEAVEMVDGVDVDRAGEIAAALDD 466
Db 672 TVSETERSLSDTAGRIAPEDTGVQELSNMDMDQAAASVSDVTTAVGDVAALGEETATEAS 731
QY 467 IADATDQQVTVVEV 481
Db 732 TRDAAAEQATTLSDV 746
RESULT 14
S55299
A:Species: Natronobacterium pharaonis
C:Species: Natronobacterium pharaonis
C:Date: 14-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 02-Sep-2000
C:Accession: S55299
R:Seidel, R.; Scharf, B.; Gautel, M.; Kleine, K.; Oesterhelt, D.; Engelhard, M.
Proc. Natl. Acad. Sci. U.S.A. 92, 3036-3040, 1995
A:Title: The primary structure of sensory rhodopsin II: a member of an additional retina
A:Reference number: S55296; MUID:95224074; PMID:7708770
A:Accession: S55299
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-534 <SEI>
A:Cross-references: EMBL:Z35086; NID:g510868; PID:g510869
A:Experimental source: strain SP1
A:Note: it is uncertain whether Met-1 or Met-22 is the initiator
C:Genetics:
A:Gene: htrII
C:Superfamily: Halobacterium salinarum transducer protein htrI
C:Keywords: signal transduction; transmembrane protein
F:22-42/Domain: transmembrane #status predicted <TM1>
F:59-79/Domain: transmembrane #status predicted <TM2>
Query Match 22.5%; Score 537.5; DB 2; Length 534;
Best Local Similarity 31.2%; Pred. No. 9.7e-17;
Matches 150; Conservative 95; Mismatches 177; Indels 59; Gaps 12;
QY 16 GIDGHALADRIGLDEA-----EIAWRLSFTGIDDDTMAALAAEQPLFEATADALVDFY-- 69
Db 72 GINLGLVAATLGGTQASLSLTAAKASRMG-DGDLDEL-----ETRDEIGDLIYAA 123
QY 70 -DHLESVERTQDLFANSTKTKVQLKETQAEYLLGLGRGEYDTEYAAQARATGKTHDVGL 128
Db 124 FDEMRSQVRLTSLDAKNAREDAEQAKRAE-----EINTELQAEAREFGEVMDRCAD 175
QY 129 GPDVYLGAITYTYTGLLDALADDVVADRGEEAATAVDELVARFLPMLKLLTFDQQAAMD 188
Db 176 GD-----FQRLDAETDN-----FAMQSIEG-----SFNEW--MDG 204
QY 189 YIDSVQAQRHLDEIDSRQELANAVATHVEAPLSLSLEATSQDVAERTDTTMRARTDDQVDRMA 248
Db 205 -IEALVGRIERFADAVSDAEAVRANAE---SVMEA-SEDNVAVQNISDRAAGDQTEVQ 259

QY 249 DVREISSVSASVEEASTADDVRRRTSEDAEALAOQGEAADDALATMTDIDEATDGVTA 308
Db 260 QIALEMDVDVATTVEEVAASADDIAKTRQAETGEAGRETAETATTEMNEVESRTEQAVA 319
QY 309 GVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEARAGEAGEGFVAVADEVKALAE 368
Db 320 SMEELNEDVREIGEVSEMIADIAEQTNILALNASIEARADGNSEGFAVVADEVKALAE 379
QY 369 SREOSTRVELVQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQETTDADVEAASG 428
Db 380 TKAATEEIDDLIGTVQDRTQTTVDIIRTSQVSEGVETVTDALERIIVDSVERTNDG 439
QY 429 MOEVSTATDEQAVSTEEVAEMVDGVDVDRAGISAAALDDIADATDQQVTVVEEVRTVKL 488
Db 440 IQEINQSIDAQAADAAQAATTWVEDMAATSEQTASDAETAETETQAEVSKEVFOLIDL 499
QY 489 S 489
Db 500 S 500
RESULT 15
A84252
Htr15 transducer [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84252
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: A84252
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-636 <STO>
A:Cross-references: GB:AE004437; NID:gi0580513; PIDN:AA019381.1; GSPDB:GN00138
C:Genetics:
A:Gene: htr15
Query Match 22.4%; Score 535.5; DB 2; Length 636;
Best Local Similarity 29.7%; Pred. No. 1.5e-16;
Matches 149; Conservative 99; Mismatches 191; Indels 63; Gaps 11;
QY 17 IDGHALADRIGLDEAETIAWRLSFTGIDDDTMAALAAEQPLFEAT---ADALVTD----- 67
Db 165 VDDDALDLSIGMP---VFMLDATG-----SVVAVNETIEELTGCSKADAMGMHASEA 214
QY 68 FYDHLSEYERTQDLFANSTKTKVQLKETQAEYLLGLGRGEYDTEYAAQARATGKTHDVGL 127
Db 215 FYDPRRVKTLAD-----KVIESPRSA--EFDIERSEKAQLADTSMV- 258
QY 128 LGPDVYLGAITYTYTGLLDALADD-----VVADRGEA--AAAVDELVARFLPML- 175
Db 259 ----TDGGTDRHIFKASPIFDDGCELLAAVEATIEDRTEDVRRADAVEELVDELSTTID 314
QY 176 -----KLLTFDQQAAMDYIDSQAQRHLDEIDSRQELANAVATHVEAPLSLEATSQ 227
Db 315 ALSSGQLSKRASFEHEGIIINEQLVSVVVALNGMADQPERLVGQVDGDTQELADTIERATA 374
QY 228 DVARTDTMRATDDQVDRMADYSREISSVSASVEEASTADDVRRRTSEDAEALAOQGEA 287
Db 375 DATDIADTV-----SSQNEMLSSAANEMENFSAHQEVAASDDQVSAAEQAQDAESGLE 430
QY 288 AADDALATMTDIDBATDGTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAAR 347
Db 431 ASEGANQATNEVIDISDDLMEVSKLESRMDEIEDVVEVIAEAVADQTNLLALNANIEAAR 490
QY 348 AGEAGEGFVAVADVVKALAEBSREQSTRVELVQMQAETEETVDQLDEVNQRIGEGVER 407

Db 491 AGEAGSGFAVVADEVKELANETREHTERIAGSISDVQQQANETVLAVEESHEQIHRAGDE 550
Qy 408 VEEAMETLQEIITDAVEDAASGMQEVSTATDEQAVSTEERVAEMVGDGVDDRAGEIAAALDDI 467
Db 551 IDDALTALEEIATSVDEAATGITEVARANDEQASTVEDVITIEDVQQQAEAAAAASDRI 610
Qy 468 ADATDQQRVTEEVRETVGKLS 489
Db 611 VSATQEQSTAVSQLSERVDKLT 632

Search completed: August 10, 2004, 15:29:26
Job time : 22.3123 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:19:00 ; Search time 15.323 Seconds
(without alignments)
1654.229 Million cell updates/sec

Title: US-09-455-978B-2

Perfect score: 2394

Sequence: 1 MSNDDTLVTADVVRGIDGH.....ATDQQVTVBEVRETVGKLS 489

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2394	100.0	489	1	HMAT_HALSA
2	2360	98.6	489	1	HMAT_HALN1
3	643	26.9	481	1	HTR3_HALN1
4	643	26.9	481	1	HTR3_HALSA
5	578	24.1	778	1	HTR6_HALN1
6	573	23.9	778	1	HTR6_HALSA
7	555	23.2	810	1	HTR4_HALN1
8	554	23.1	810	1	HTR4_HALSA
9	537.5	22.5	534	1	HTR2_NATPH
10	521	21.8	545	1	HTR5_HALN1
11	492	20.6	433	1	HTR2_HALVA
12	464	19.4	763	1	HTR2_HALN1
13	463.5	19.4	535	1	HTR1_HALN1
14	461	19.3	764	1	HTR2_HALSA
15	449.5	18.8	535	1	HTR1_HALSA
16	412.5	17.2	432	1	HMAT_BACSU
17	388.5	16.2	662	1	MCPB_BACSU
18	386.5	16.1	682	1	PLIJ_PSEAF
19	379	15.8	662	1	TLPB_BACSU
20	359.5	15.0	547	1	MCPC_SALTY
21	359.5	15.0	1000	1	Y041_SYNV3
22	354.5	14.8	662	1	TLPB_BACSU
23	351	14.7	548	1	HLVB_VIBCH
24	349	14.6	553	1	MCP2_ECOLI
25	348.5	14.6	551	1	MCP1_ECOLI
26	347.5	14.5	573	1	TUPC_BACSU
27	342.5	14.3	657	1	MCPA_CAUCR
28	335.5	14.0	553	1	MCP2_SALTY
29	334	14.0	661	1	MCPA_BACSU
30	333.5	13.9	654	1	MCP3_BACSU
31	326.5	13.6	546	1	MCP3_ECOLI
32	323.5	13.5	668	1	DCRA_DESVH
33	318.5	13.3	557	1	MCP5_ENTAE

RESULT 1

ID	HMAT_HALSA	STANDARD;	PRT;	489 AA.
AC	P71413			
DT	28-FEB-2003	(Rel. 41, Created)		
DT	28-FEB-2003	(Rel. 41, Last sequence update)		
DT	28-FEB-2003	(Rel. 41, Last annotation update)		
DE	Heme-based aerotactic transducer hemAR (Transducer HtB protein).			
CN	HEMAT OR HTB.			
OS	Halobacterium salinarum.			
OC	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;			
OC	Halobacteriaceae; Halobacterium.			
OX	NCBI_TaxID=2242;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FLXJ5.			
RX	MEDLINE=96209786; PubMed=8643458;			
RA	Zhang W., Brooun A., McCandless J., Banda P., Alam M.;			
RT	"Signal transduction in the archaeon Halobacterium salinarum is processed through three subfamilies of 13 soluble and membrane-bound transducer proteins.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).			
RN	[2]			
RP	FUNCTION.			
RX	MEDLINE=20140131; PubMed=10676961;			
RA	Hou S., Larsen R.W., Boudko D., Riley C.W., Karatan E., Zimmer M.,			
RT	Ordal G.W., Alam M.;			
RL	"Myoglobin-like aerotaxis transducers in Archaea and Bacteria.";			
CC	Nature 403:540-544(2000).			
CC	-1- FUNCTION: Heme-containing signal transducer responsible for aerotaxis, the migratory response toward or away from oxygen.			
CC	-1- SUBUNIT: Homotetramer (Probable).			
CC	-1- SIMILARITY: Contains 1 methyl-accepting transducer domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; U75436; AAB17881.1; -.			
DR	PIR; T44978; T44978.			
DR	HSSP; P02942; 10U7.			
DR	GO; GO:0020037; F:heme binding; IDA.			
DR	GO; GO:0004871; F:signal transducer activity; IDA.			
DR	GO; GO:0009454; P:aerotaxis; IDA.			
DR	GO; GO:0007165; P:signal transduction; IDA.			
DR	InterPro; IPR004089; Chmtaxis trans.			
DR	InterPro; IPR004090; Me chemotaxis.			
DR	Pfam; PF00015; MCPsignal; 1.			
DR	PRINTS; PR00260; CHEMTNSDUCR.			
DR	SMART; SM00283; MA; 1.			
DR	PROSITE; PS0111; CHEMOTAXIS_TRANSDUC_2; 1.			
KW	Transducer; Heme.			

Q52877 rhizobium m
Q02929 clostridium
P21823 enterobacte
P55652 rhizobium s
P50466 escherichia
P07018 escherichia
P55439 rhizobium s
P29486 vibrio chol
P43000 myxococcus
P46865 leishmania
P35749 homo sapien
P35748 oryctolagus

ALIGNMENTS

ID	HMAT_HALSA	STANDARD;	PRT;	489 AA.
AC	P71413			
DT	28-FEB-2003	(Rel. 41, Created)		
DT	28-FEB-2003	(Rel. 41, Last sequence update)		
DT	28-FEB-2003	(Rel. 41, Last annotation update)		
DE	Heme-based aerotactic transducer hemAR (Transducer HtB protein).			
CN	HEMAT OR HTB.			
OS	Halobacterium salinarum.			
OC	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;			
OC	Halobacteriaceae; Halobacterium.			
OX	NCBI_TaxID=2242;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FLXJ5.			
RX	MEDLINE=96209786; PubMed=8643458;			
RA	Zhang W., Brooun A., McCandless J., Banda P., Alam M.;			
RT	"Signal transduction in the archaeon Halobacterium salinarum is processed through three subfamilies of 13 soluble and membrane-bound transducer proteins.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).			
RN	[2]			
RP	FUNCTION.			
RX	MEDLINE=20140131; PubMed=10676961;			
RA	Hou S., Larsen R.W., Boudko D., Riley C.W., Karatan E., Zimmer M.,			
RT	Ordal G.W., Alam M.;			
RL	"Myoglobin-like aerotaxis transducers in Archaea and Bacteria.";			
CC	Nature 403:540-544(2000).			
CC	-1- FUNCTION: Heme-containing signal transducer responsible for aerotaxis, the migratory response toward or away from oxygen.			
CC	-1- SUBUNIT: Homotetramer (Probable).			
CC	-1- SIMILARITY: Contains 1 methyl-accepting transducer domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; U75436; AAB17881.1; -.			
DR	PIR; T44978; T44978.			
DR	HSSP; P02942; 10U7.			
DR	GO; GO:0020037; F:heme binding; IDA.			
DR	GO; GO:0004871; F:signal transducer activity; IDA.			
DR	GO; GO:0009454; P:aerotaxis; IDA.			
DR	GO; GO:0007165; P:signal transduction; IDA.			
DR	InterPro; IPR004089; Chmtaxis trans.			
DR	InterPro; IPR004090; Me chemotaxis.			
DR	Pfam; PF00015; MCPsignal; 1.			
DR	PRINTS; PR00260; CHEMTNSDUCR.			
DR	SMART; SM00283; MA; 1.			
DR	PROSITE; PS0111; CHEMOTAXIS_TRANSDUC_2; 1.			
KW	Transducer; Heme.			

```

FT DOMAIN 218 454 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 489 AA; 52816 MW; 8457263FCF616BFF CRC64;

Query Match
Best Local Similarity 100.0%; Score 2394; DB 1; Length 489;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNDNDTLVADVNRNGIDGHADRLGDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
DB 1 MSNDNDTLVADVNRNGIDGHADRLGDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
QY 61 ADALVTFDYHLESYERTQDLFANSTKTVQEQKTAQAEVLLGLGRGEYDTYAAQARIG 120
DB 61 ADALVTFDYHLESYERTQDLFANSTKTVQEQKTAQAEVLLGLGRGEYDTYAAQARIG 120
QY 121 KIHVDVLGDPVYLGAYTRYTGLLDALADDVADRGEAAAAVDELVARFLPKLLTF 180
DB 121 KIHVDVLGDPVYLGAYTRYTGLLDALADDVADRGEAAAAVDELVARFLPKLLTF 180
QY 181 DQIAMDYIDSVAORLHDEIDSRQELANAVATHVEAPLSLEATSDVAERTDTMRAAT 240
DB 181 DQIAMDYIDSVAORLHDEIDSRQELANAVATHVEAPLSLEATSDVAERTDTMRAAT 240
QY 241 DDQVDRMADVSREISSVSASVEEASTADVVRTSDEAALAQGEAAADALATMTDID 300
DB 241 DDQVDRMADVSREISSVSASVEEASTADVVRTSDEAALAQGEAAADALATMTDID 300
QY 301 EATDGTAGVQELGERAADVESVTGIDDIAGQTNMLNALSIEAARAGEGFAVVAD 360
DB 301 EATDGTAGVQELGERAADVESVTGIDDIAGQTNMLNALSIEAARAGEGFAVVAD 360
QY 361 EVKALAEESRQSTRVEELVEQMAETETVDQLDEVNQRIGGVVERVEEAMETLQETD 420
DB 361 EVKALAEESRQSTRVEELVEQMAETETVDQLDEVNQRIGGVVERVEEAMETLQETD 420
QY 421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVGVDDRAGEIAALDDIADATDQOVRTVEE 480
DB 421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVGVDDRAGEIAALDDIADATDQOVRTVEE 480
QY 481 VRETGVKLS 489
DB 481 VRETGVKLS 489

RESULT 2
ID HMAT HALN1 STANDARD; PRT; 489 AA.
AC QSHPR5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Heme-based aerotactic transducer hemaT.
GN HEMAT OR HTR10 OR VNG1505G.
OC Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahaitas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weiti R., Goo Y.A.,
RA Leitthausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Ikenbarger T.A., Peck R.F., Pohlischer M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- FUNCTION: Heme-containing signal transducer responsible for
CC aerotaxis, the migratory response toward or away from oxygen (By
CC similarity).

```

```

CC -!- SUBUNIT: Homotetramer (Probable).
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AE005064; AAG19801.1; -
DR PIR; E84304; E84304.
DR HSP; P02942; I007.
DR InterPro; IPR004089; Chmtaxis transd.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTNSDUCR.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS00111; CHEMOTAXIS_TRANSDUC_2; 1.
KW Transducer; Heme; Complete proteome.
FT DOMAIN 218 454 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 489 AA; 52852 MW; 8C0AEF179667791E CRC64;

Query Match
Best Local Similarity 98.6%; Score 2360; DB 1; Length 489;
Matches 484; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSNDNDTLVADVNRNGIDGHADRLGDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
DB 1 MSNDNDTLVADVNRNGIDGHADRLGDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
QY 61 ADALVTFDYHLESYERTQDLFANSTKTVQEQKTAQAEVLLGLGRGEYDTYAAQARIG 120
DB 61 ADALVTFDYHLESYERTQDLFANSTKTVQEQKTAQAEVLLGLGRGEYDTYAAQARIG 120
QY 121 KIHVDVLGDPVYLGAYTRYTGLLDALADDVADRGEAAAAVDELVARFLPKLLTF 180
DB 121 KIHVDVLGDPVYLGAYTRYTGLLDALADDVADRGEAAAAVDELVARFLPKLLTF 180
QY 181 DQIAMDYIDSVAORLHDEIDSRQELANAVATHVEAPLSLEATSDVAERTDTMRAAT 240
DB 181 DQIAMDYIDSVAORLHDEIDSRQELANAVATHVEAPLSLEATSDVAERTDTMRAAT 240
QY 241 DDQVDRMADVSREISSVSASVEEASTADVVRTSDEAALAQGEAAADALATMTDID 300
DB 241 DDQVDRMADVSREISSVSASVEEASTADVVRTSDEAALAQGEAAADALATMTDID 300
QY 301 EATDGTAGVQELGERAADVESVTGIDDIAGQTNMLNALSIEAARAGEGFAVVAD 360
DB 301 EATDGTAGVQELGERAADVESVTGIDDIAGQTNMLNALSIEAARAGEGFAVVAD 360
QY 361 EVKALAEESRQSTRVEELVEQMAETETVDQLDEVNQRIGGVVERVEEAMETLQETD 420
DB 361 EVKALAEESRQSTRVEELVEQMAETETVDQLDEVNQRIGGVVERVEEAMETLQETD 420
QY 421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVGVDDRAGEIAALDDIADATDQOVRTVEE 480
DB 421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVGVDDRAGEIAALDDIADATDQOVRTVEE 480
QY 481 VRETGVKLS 489
DB 481 VRETGVKLS 489

RESULT 3
ID HTR3 HALN1 STANDARD; PRT; 481 AA.
AC Q9H000;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Halobacterial transducer protein III.

```



```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Halobacterial transducer protein VI.
GN HTR4 OR HTPV1.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S9;
RX MEDLINE=96275896; PubMed=8674984;
RA Rudolf J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RA Oesterhelmt D.;
RT "A family of halobacterial transducer proteins.";
RL FEMS Microbiol. Lett. 139:161-168 (1996).
CC -!- FUNCTION: Potentially involved in chemo- or phototactic signal
CC transduction.
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC -!- SIMILARITY: Contains 2 HAMP domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DE EMBL; X95590; CAA64843.1; -.
DR PIR; T48897; T48897.
DR HSSP; P02942; 1QU7.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_Chemotaxis.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 2.
DR Transducer; Transmembrane; Repeat.
KW DOMAIN 1 26 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 27 47 POTENTIAL.
FT DOMAIN 48 296 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 297 317 POTENTIAL.
FT DOMAIN 318 778 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 318 370 HAMP 1.
FT DOMAIN 412 465 HAMP 2.
FT DOMAIN 484 720 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 778 AA; 82077 MW; 134C7D7F0A3334CD CRC64;

Query Match 23.9%; Score 573; DB 1; Length 778;
Best Local Similarity 33.0%; Pred. No. 1.3e-17;
Matches 144; Conservative 74; Mismatches 168; Indels 50; Gaps 6;

QY 88 TVEQLKETCAEYLLGLGRGEYDTEYAAQRA-RIGKIHVDVLGPDVYLGAITYTYTGILD 146
DB 320 TVRALNDLEAK-AAALERGEYDITLDVARVDLGRLEFAFASLRDTVQARINDANEQQVD 378
QY 147 -----ALADVDVADRGEAAAVDELVAREL-----PMLKLLTFDQOIAMD 187
DB 379 AEAARSEAAQADAEAAQAEAAEAESEAAQARLETTAEAFSETMAYAYAGDLTVRLD 438
QY 188 TVIDSYAQ-----RUHIDSRQELANAVATHVEAPLSLEAT 225
DB 439 ADVEQAAMADIAAFENWAAADMEATIADAGRADEVAT-----ASTDASDAAAVEOT 491
QY 226 SDVAERTDTWEARTDDQVDRADYVSRELSSVSASVEEVASTADDVRITSEDAEALAAQQ 285
DB 492 GHDVSDAVGRIDRAADQRDQLEAVASETDEMSATIEEVAASDAQVETSQRAAALGDG 551
QY 286 EAAADDALATMTDIDEATDGTAGVQLGERAADVESVTGVIDDIADAEQTNMLALNASTEA 345

Db 552 QAAQDAVAQAEEIEDETEQAAATAVDDLEAKSEIETIIVAALTDAEQTNMLALNASTEA 611
QY 346 ARAGEAGEGFVAVADEVKALAESEREQSTRVEELVEQMAETETVVDQLDEVNQRIGEGV 405
Db 612 ARADQDGDGFVAVADEVKDLADESKASAEIEALVAEVAQVETSVAAAMDRIQERVSDGV 671
QY 406 ERVEEAMETLQBITDAVEDAASGMQEVSTATDEQAVSTEVEAEMVDGVDDRAGEIAAALD 465
Db 672 ETVSETERSLSIAGRIAEADTVGOEISNAMDDQAASVSDVTTAVGDVAALGEETATEAE 731
QY 466 DIADATDQOQVTVVEV 481
Db 732 STADAAAEQASTLSDV 747

RESULT 7
HTR4_HALN1
ID_HTR4_HALN1 STANDARD; PRT; 810 AA.
AC Q9HP84;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Halobacterial transducer protein IV.
GN HTR5 OR HTPIV OR VNG1760G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA NG W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Pahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RA "Genome sequence of Halobacterium species NRC-1.";
RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
CC -!- FUNCTION: Potentially involved in chemo- or phototactic signal
CC transduction (By similarity)
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC -!- SIMILARITY: Contains 2 HAMP domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DE EMBL; AE005080; AAG19986.1; -.
DR PIR; F84327; F84327.
DR HSSP; P02942; 1QU7.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 2.
KW Transducer; Transmembrane; Complete proteome; Repeat.
FT DOMAIN 1 38 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 39 59 POTENTIAL.
FT DOMAIN 60 323 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 324 344 POTENTIAL.
FT DOMAIN 345 810 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 345 397 HAMP 1.

```

```

FT DOMAIN 439 493 HAMP 2.
FT DOMAIN 512 748 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 810 AA; 85219 MW; 4BF36E4B7D22B80 CRC64;

Query Match 23.2%; Score 555; DB 1; Length 810;
Best Local Similarity 32.1%; Pred. No. 7.5e-17;
Matches 142; Conservative 85; Mismatches 174; Indels 42; Gaps 8;

QY 84 NTKTKVQLKETQAEVLLGLGRGYDTEVAQR-----ARIGKIHVILGPDVYLG 135
DB 346 NTAATAVQSLSAARAE-----IEAGNYDQVAVSSRRDEIGQLFASIGSMRDALVTQIDEAEA 401
QY 136 ATRYTYTGLLDALAD-----DVVADRCGEAAAVDELVA-----RFLPMLKLL----- 178
DB 402 AREQATEAQDAAERERAEADAKADAALAELEAQAERYSDVMAACADGDLT 461
QY 179 -----TPDQOIA--MDTYIDSYAQLHDEIDSROELANAVAT---HVEAPLSSLEATS 226
DB 462 RMPADDTTNEAMAALAAAFNEMLAQWEHTIIDI-QEFADAVATASEEAEGVGAADAEAS 520
QY 227 QVVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVVRTSEDALAAQOGE 286
DB 521 GQVSESVQEIAGADEQRNMLDTVSGEMTDLSAAIEEVAASADSVAESHQTAETIARDGE 580
QY 287 AAADALATMTDIDEATDGTAGVQLGERADVESVTGVIDDIAEQTMLNALNASTIEAA 346
DB 581 QTAEDAERSLSVQEAIDATVQVNEALDDQMAEISEIVDLISDIAEQTMLNALNANIEAA 640
QY 347 RAGEAGEGFVAVADEVKALAEESREQSTRVEELVEQMAETETVDQLDEVNQRIEGEVE 406
DB 641 RADKSGDGFVAVADEVKOLAEETQESAGDIERRITEVQSQTATVAEAAEESMDAGID 700
QY 407 RVEEAMETLOEITDAVEDAASQGVSTATDQAVSTEEVAEMVQGVDDRAGEIAALDD 466
DB 701 AVEEVVDAFTAVSDHADETDGTQVEISDTTDDQASSTEEAVSMTEEVAADLSSTAGEAQS 760
QY 467 IADATDQQVTVTEEVRETGKLS 489
DB 761 VSAAAEEQAASMSSEISDSVESLS 783

RESULT 8
ID HTR4 HALSA STANDARD; PRT; 810 AA.
AC Q48317;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Halobacterial transducer protein IV.
GN HTR5 OR HTPIV.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S9;
RX MEDLINE=96275896; PubMed=8674984;
RA Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RA Oesterheilt D.;
RT "A family of halobacterial transducer proteins.";
RL FEMS Microbiol. Lett. 139:161-168(1996).
CC -!- FUNCTION: Potentially involved in chemo- or phototactic signal transduction.
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC -!- SIMILARITY: Contains 2 HAMP domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X95589; CAAG4841.1; -.
DR PIR; T46810; T46810.
DR HSSP; P02942; LQUT7.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS00885; HAMP; 2.
KW Transducer; Transmembrane; Repeat.
FT DOMAIN 1 38 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 39 59 POTENTIAL.
FT DOMAIN 60 323 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 324 344 POTENTIAL.
FT DOMAIN 345 810 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 345 397 HAMP 1.
FT DOMAIN 439 493 HAMP 2.
FT DOMAIN 512 748 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 810 AA; 85207 MW; 37B0F6046A39D9BA CRC64;

Query Match 23.1%; Score 554; DB 1; Length 810;
Best Local Similarity 32.1%; Pred. No. 8.3e-17;
Matches 142; Conservative 85; Mismatches 174; Indels 42; Gaps 8;

QY 84 NTKTKVQLKETQAEVLLGLGRGYDTEVAQR-----ARIGKIHVILGPDVYLG 135
DB 346 NTAATAVQSLSAARAE-----IEAGNYDQVAVSSRRDEIGQLFASIGSMRDALVTQIDEAEA 401
QY 136 ATRYTYTGLLDALAD-----DVVADRCGEAAAVDELVA-----RFLPMLKLL----- 178
DB 402 AREQATEAQDAAERERAEADAKADAALAELEAQAERYSDVMAACADGDLT 461
QY 179 -----TPDQOIA--MDTYIDSYAQLHDEIDSROELANAVAT---HVEAPLSSLEATS 226
DB 462 RMPADDTTNEAMAALAAAFNEMLAQWEHTIIDI-QEFADAVATASEEAEGVGAADAEAS 520
QY 227 QVVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVVRTSEDALAAQOGE 286
DB 521 GQVSESVQEIAGADEQRNMLDTVSGEMTDLSAAIEEVAASADSVAESHQTAETIARDGE 580
QY 287 AAADALATMTDIDEATDGTAGVQLGERADVESVTGVIDDIAEQTMLNALNASTIEAA 346
DB 581 QTAEDAERSLSVQEAIDATVQVNEALDDQMAEISEIVDLISDIAEQTMLNALNANIEAA 640
QY 347 RAGEAGEGFVAVADEVKALAEESREQSTRVEELVEQMAETETVDQLDEVNQRIEGEVE 406
DB 641 RADKSGDGFVAVADEVKOLAEETQESAGDIERRITEVQSQTATVAEAAEESMDAGID 700
QY 407 RVEEAMETLOEITDAVEDAASQGVSTATDQAVSTEEVAEMVQGVDDRAGEIAALDD 466
DB 701 AVEEVVDAFTAVSDHADETDGTQVEISDTTDDQASSTEEAVSMTEEVAADLSSTAGEAQS 760
QY 467 IADATDQQVTVTEEVRETGKLS 489
DB 761 VSAAAEEQAASMSSEISDSVESLS 783

```

```

RESULT 9
ID HTR2 NATPH STANDARD; PRT; 534 AA.
AC P42259;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis
DE protein II) (MPP-II).
GN HTR2 OR HTRII.
OC Natronomonas pharaonis (Natronobacterium pharaonis).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;

```


CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE005080; AAG19985.1; --
 DR EMBL; X95589; CAA64842.1; --
 DR PIR; E84327; E84327.
 DR FIR; T46811; T46811.
 DR HSP; P02942; 1Q07.
 DR InterPro; IPR004089; Chmtaxis_trans.
 DR InterPro; IPR003660; HAMP.
 DR InterPro; IPR004090; Me.chemotaxis.
 DR Pfam; PF00015; MCPsignal; 1.
 DR PRINTS; PR00260; CHEMTRNSDUCR.
 DR SMART; SM00304; HAMP; 1.
 DR SMART; SM00283; MA; 1.
 DR PROSITE; P50111; CHEMOTAXIS_TRANSDUC_2; 1.
 DR PROSITE; P50885; HAMP; 1.
 KW Transducer; Transmembrane; Complete proteome.
 FT TRANSMEM 10 30 POTENTIAL.
 FT TRANSMEM 44 64 POTENTIAL.
 FT TRANSMEM 80 100 POTENTIAL.
 FT DOMAIN 164 217 HAMP.
 FT DOMAIN 236 476 METHYL-ACCEPTING TRANSDUCER.
 SQ SEQUENCE 545 AA; 57070 MW; FDD870389C2F428B CRC64;

Query Match 21.8%; Score 521; DB 1; Length 545;

Best Local Similarity 29.6%; Pred. No. 1.3e-15;

Matches 142; Conservative 94; Mismatches 195; Indels 48; Gaps 10;

QY 28 LDEAETAWLSFTGIDDDTMAALAAEQPLFEATADALVTDF----YDHLESYERTQDLFA 93
 DB 64 LEKAGVATQI--FNLTGPTLMTAVVFAIGGTATLAIEDMEALVEEARQAEAEERA 122
 QY 84 NSTKTVEQLKETQAEYLLGLGRGEYDEYAARIGIKIHVDVLGLGPDVVLGAYTRYTG 143
 DB 123 EAERAREKAQKQAE----AERQTAEASAKQDAR--ERSAEIEQLAAD--LESQATEVGA 175
 QY 144 LLDALAD-----DVVARGE--EAAAADVELVARFLPMLKLLTFDQIAMDYIDSVA 194
 DB 176 TLEAASDGLTARVDATTDNAETAEVATVNDM-----LTTMERTI----- 216
 QY 195 QRLHDEIDSRQELANAVATHVEAPLSLEATSDQVAERTDTMRAPDDQVDRMADYSRI 254
 DB 217 ----DSIQGFSTNTVTASREATAKKEIQDASQTVSESVOEIAAGTDDQREQLSVAEM 272
 QY 255 SVSASVEEVAETADVDVRTSDEAALAQGGAAADALATMTDIDEATDGTAGVEQLG 314
 DB 273 DSYSATVEEVAATAQSVADTAADTTDVATAGTAEADAIDAIDAVQETMTQTVANVDAL 332
 QY 315 ERAADVESVTGIDDIATQTNMLNASTAEARAGEA-----GEGFAVVADEKALAEER 370
 DB 333 DUTTRIDDAELISDTAEQTNMLNANAEARAGSGGSGSDGFAVVADEKALATESQ 392
 QY 371 EOSTVEELVEOMQAEETETVQDLDRVNRIGEGVERVEEAMETLQEIITDAVEDAASGMQ 430
 DB 393 RSAKDIAELIEVQSQTATTVREIVAEORVNDGAAAEVETVDAFGAVTENIQETDVGQ 452
 QY 431 EYSTATDEQAVSTEVEAEVMDGVDRAAALDDIADATDQVQRTVEEVRTVGKLS 489
 DB 453 EISQAMDEQAQSRVSVSSVDDIATISQATADRAENVSAASEQTASITEVTSLSQLA 511

RESULT 11

ID HTR2 HALVA
 AC P42258; STANDARD; PRT; 433 AA.
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis

DE protein II) (MPP-II) (Fragment).
 GN HTRII.
 OS Haloarcula vallismortis.
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Haloarcula.
 OX NCBI_TaxID=28442;
 RN [1]_TaxID=28442;
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29715;
 RX MEDLINE=95224074; PubMed=7708770;
 RA Seidel R., Scharf B., Gautel M., Kleine K., Oesterhelt D.,
 RT Engelhard M.;
 RT "The primary structure of sensory rhodopsin II: a member of an
 RT additional retinal protein subgroup is coexpressed with its
 RT transducer, the halobacterial transducer of rhodopsin II.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3036-3040(1995).
 CC -!- FUNCTION: Transduces signals from the phototaxis receptor sensory
 CC rhodopsin II (SR-II) to the flagellar motor. Responds to light
 CC changes through the variation of the level of methylation. Also
 CC acts as a chemotransducer.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
 CC -!- SIMILARITY: Contains 1 HAMP domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Z35308; CAA84549.1; --
 DR HSP; P02942; 1Q07.
 DR InterPro; IPR004089; Chmtaxis_trans.
 DR InterPro; IPR003660; HAMP.
 DR InterPro; IPR004090; Me.chemotaxis.
 DR Pfam; PF00015; MCPsignal; 1.
 DR PRINTS; PR00260; CHEMTRNSDUCR.
 DR SMART; SM00304; HAMP; 1.
 DR SMART; SM00283; MA; 1.
 DR PROSITE; P50111; CHEMOTAXIS_TRANSDUC_2; 1.
 DR PROSITE; P50885; HAMP; 1.
 KW Transducer; Photoreceptor; Chemotaxis; Transmembrane; Methylation.
 FT NON_TER 1 1
 FT DOMAIN 58 111 HAMP.
 FT DOMAIN 130 366 METHYL-ACCEPTING TRANSDUCER.
 SQ SEQUENCE 433 AA; 45935 MW; 90507B8897D943C0 CRC64;

Query Match 20.6%; Score 492; DB 1; Length 433;
 Best Local Similarity 29.9%; Pred. No. 1.8e-14;
 Matches 127; Conservative 86; Mismatches 172; Indels 40; Gaps 5;

QY 69 YDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDEYAARIGIKIHVDLGL 128
 DB 11 FDNMRANLRTQISEATAKQAEAAKEQAQ-----AAREDVESERNEMELTG----- 58
 QY 129 GPDVVLGAYTRYTGLLDALADD-----VADRGEEAAAADVELVARFLPMLKLLTFDQ 183
 DB 59 ----HLELKAQOQYSDALDAANGDLTARVKTDSMDMAEVEGDIINTTLDALDIEDTADM 114
 QY 184 IAMDVTYDSYAQRHLDEIDSRQELANAVATHVEAPLSLEATSDQVAERTDTMRAPDDQ 243
 DB 115 AFATNVIQS-----SDRVNSNAE-----RVDRASKQVSKSINEIFSGTTEQ 155
 QY 244 VDRMADVSRRTSSVSAEVEEVAETADVDVRTSDEAALAQGGAAADALATMTDIDEAT 303
 DB 156 NEGLESAAAEMQNLGATAQVASSAQVADTSQSAKVGEDGREAQAIAEWSAIEART 215
 QY 304 DGVTAGVQLGERAADVESVTGIDDIATQTNMLNASTAEARAGEAEGEFAVVADEVK 363
 DB 216 GETVEEINALDDELDEIGEIVGVITTSIVEQTNMLNASTAEAAHADGEGEFAVVADEIK 275

```

QY 364 ALAESREQSTRVEELVEQMAETEETVQDQVNDQVNRIGGEVVERVEAMETLQEIITDAVE 423
Db 276 GLAEETKEAADIIEGRALQAGQVETVETMESTSTRITEGYSTVETVDALETIVEYTE 335
QY 424 DRASGMQEVSTATDEQAVSTEEVAEMVGDVDRAGISAAALDDIADATDQQVTRVEEVR 483
Db 336 EVDVTGIEIDRATEEQARTAQDVMTGIDTIDTTISQOTATEADTVAGAAQDQASIEEVS 395
QY 484 TVGKL 488
Db 396 SATEL 400

RESULT 12
HTR2 HALN1
ID HTR2 HALN1 STANDARD; PRT; 763 AA.
AC Q9HF61;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis
DE protein II) (MPP-II).
GN HTR2 OR VNGI765G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- FUNCTION: Transduces signals from the phototaxis receptor sensory
CC rhodopsin II (SR-II) to the flagellar motor. Responds to light
CC changes through the variation of the level of methylation. Also
CC acts as a chemotransducer (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC -!- SIMILARITY: Contains 2 HAMP domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005080; AAG19989.1; -.
DR PIR; A84328; A84328.
DR HSP; P02942; LQ07.
DR InterPro; IPR004089; Chntaxis_transd.
DR InterPro; IPR003650; HAMP.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; NA; 1.
DR PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS00885; HAMP; 2.
KW Transducer; Photoreceptor; Chemotaxis; Transmembrane; Methylation;
KW Complete proteome; Repeat.
FT INIT_MET 0 0 BY SIMILARITY.
FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 16 36 POTENTIAL.
FT DOMAIN 37 277 EXTRACELLULAR (POTENTIAL).

```

```

FT TRANSMEM 278 297 POTENTIAL.
FT DOMAIN 298 763 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 302 354 HAMP 1.
FT DOMAIN 396 449 HAMP 2.
FT DOMAIN 468 704 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 763 AA; 78911 MW; CF7A8FF04DFF309A CRC64;

Query Match 19.4%; Score 464; DB 1; Length 763;
Best Local Similarity 28.9%; Pred. No. 4.8e-13;
Matches 136; Conservative 74; Mismatches 219; Indels 42; Gaps 9;

QY 37 LSFTGIDDDTMAALAAEQPLFEATDALVTDFVHLESYERTQDLFANSTKTKVQLKETQ 96
Db 293 LSLIGTVGSTTVTALRQ--FSRRADEMAAGDLDTIDTSR-NDEGTTAEFSRMRDSL 349
QY 97 AEYLLGLGRGEYDTEYAAQARIGKIHVLGPNVYLGYTRYTYTGLLDALADVVADR 156
Db 350 SSSLT-----DAERATARAEDAR-EDAEQQRADA-----EAAEDAEAAAR 388
QY 157 --GEEAANAVDLVARFLPMLKLT---FDQIAMDYIDSVAQRLLH-----DEIDSQ 205
Db 389 KDAQETARALESAADYEALTAVDAGDLTRRVDASRDHDAMARIGHALNDMLDDIETSV 448
QY 206 ELANAVATVHEAPLGLSLEA-----TSQDVAERTDTMRATDDQVDRMADYSREISSVS 258
Db 449 AATAFSDHVSDAQORVEADAGDAIDAGTIDVSTAVDEISDGATEQDRLHVEAGEVDDLS 508
QY 259 ASVEEVASTADVTRTSEDAEALAOQGEAAADALATMTDIDEATDGVTVAGVEQLGERAA 318
Db 509 ASAEVETVSLADTAGAASAVDDGQATDEAVETMDVDADDAEAAAADAMDALDSEMA 568
QY 319 DVESVTGVDIDDAEOTNMLNASTAEARAGAGGFAVVADEVKALAESEBQSTRVEE 378
Db 569 DIGEIVDVADIADQTNMLNASTAEARTGADGDFAVVADEVKTLAESEDAEDIES 628
QY 379 LVEQMAETEETVQDQVNDQVNRIGGEVVERVEAMETLQEIITDAVEDAAGMGEVSTATDB 438
Db 629 RLALQGGQVSDVADENRATSDTVSGRATVGDAATALDDVSVFVADTDTAAQEIRANTDR 688
QY 439 QAVSTEVAEMVGDVDRAGEIAAALDDIADATDQQVTRVEEVRVTVGKLS 489
Db 689 QAHAASRVASAVDEVAGISQETAQATAVADSAATQTDTLSSVDDAAADLA 739

RESULT 13
HTR1 HALN1
ID HTR1 HALN1 STANDARD; PRT; 535 AA.
AC P33741; Q9HPP6;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sensory rhodopsin I transducer (HTR-I) (Methyl-accepting phototaxis
DE protein I) (MPP-I).
GN HTR1 OR HTR1 OR VNGI659G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081), and
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091; 2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRC-1;
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).

```



```

RC STRAIN=L33;
RX MEDLINE=90384855; PubMed=2205842;
RA Ferrando E., Krah M., Marwan W., Oesterhelt D.;
RT "A gene from S. pombe with homology to E. coli RNase III blocks
RT conjugation and sporulation when overexpressed in wild type cells.";
RN Nucleic Acids Res. 18:5304-5304(1990).
RL [2]
RP FUNCTION.
RX MEDLINE=94244615; PubMed=8187768;
RA Krah M., Marwan W., Vermiglio A., Oesterhelt D.;
RT "Phototaxis of Halobacterium salinarum requires a signalling complex
RT of sensory rhodopsin I and its methyl-accepting transducer HtrI.";
RL EMO J. 13:2150-2155(1994).
CC -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
CC SENSORY RHODOPSIN I (SR-I) TO THE FLAGELLAR MOTOR. RESPONDS TO
CC LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC -!- SIMILARITY: Contains 2 HAMP domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68591; CAA48578.1; -.
DR HSSP; P02942; 1OU7.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_chemotaxis.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTXNSDUCR.
DR SMART; SMO0304; HAMP; 2.
DR SMART; SMO0283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 2.
KW Transducer; Photoreceptor; Transmembrane; Methylation; Repeat.
FT INIT_MET 0 0
FT DOMAIN 1 8
FT TRANSMEM 9 29
FT DOMAIN 30 33
FT TRANSMEM 34 54
FT DOMAIN 55 535
FT DOMAIN 54 106
FT DOMAIN 148 201
FT DOMAIN 220 458
FT MOD_RES 265 265
FT MOD_RES 272 272
FT MOD_RES 279 279
FT MOD_RES 463 463
FT MOD_RES 472 472
FT SEQUENCE 535 AA; 56814 MW; 504A165A47FA8A45 CRC64;
Query Match 18.8%; Score 449.5; DB 1; Length 535;
Best Local Similarity 17.7%; Pred. No. 1.4e-12;
Matches 141; Conservative 86; Mismatches 191; Indels 91; Gaps 15;
Qy 28 LDEAEIAWRLSFTGDDDTMAALAEQPLFEATADALVTFDYDHLESYERTQDLFANSTK 87
Db 51 INAGETVGRKEIGAQTERVANGNLQEVTSRTD-----EFGSLADSIQMQLSLGRGLN 106
Qy 88 TVEQ----LKETAQYLLGLGRGEYDTEVAARIGTKIHDVLGLGPVYLGAITYRYTG 143
Db 107 EMERTSRDLEETQTE-----SETTRESEQAQEQSAE-----REARE 145
Qy 144 LLDALADVVVADRG-----EAAAVDELVARFLPKLLTTFDQOI-----AMDYIDSYA 194
Db 146 LAATYQDSNRD--GETMNESSATGD-----FTQKVDVDTDPENMETGTAFN 190
Qy 195 QRLHDEIDSRQELANAVATHEAPLSSLEATSDQV--AERTDTPWA-----RTDQVDRM 247

```

Search completed: August 10, 2004, 15:26:49
Job time : 16.3923 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:22:05 ; Search time 58.609 Seconds

(without alignments)
2632.505 Million cell updates/sec

Title: US-09-455-978B-2

Perfect score: 2394

Sequence: 1 MSNDNDTLVTADVRNGIDGH.....ATDQQVTVVEYRVETGKLS 489

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Description
1	609.5	25.5	P71412 halobacteri
2	597	24.9	Q9HP10 halobacteri
3	597	24.9	Q9P9J0 halobacteri
4	580	24.2	Q9HR88 halobacteri
5	560.5	23.4	Q93775 halobacteri
6	540.5	22.6	P71415 halobacteri
7	539	22.5	P71414 halobacteri
8	535.5	22.4	Q9HQX8 halobacteri
9	529	22.1	P71416 halobacteri
10	524	21.9	O06022 halobacteri
11	523.5	21.9	P71409 halobacteri
12	523.5	21.9	Q9HPQ5 halobacteri
13	521.5	21.8	Q9HRN6 halobacteri
14	509.5	21.3	Q9HRA1 halobacteri
15	507.5	21.2	O93643 halobacteri
16	506	21.1	Q9HPW6 halobacteri

17	505	21.1	642	1	O59634
18	502.5	21.0	773	1	Q9UXS0
19	499.5	20.9	627	17	Q9HS86
20	497.5	20.8	419	1	O93644
21	497	20.8	536	17	Q9HPA2
22	475.5	19.9	537	16	Q9AB06
23	452.5	18.9	423	17	Q9HQT5
24	439.5	18.4	423	1	O93642
25	419.5	17.5	551	16	Q815U4
26	419.5	17.5	559	16	Q9K669
27	415.5	17.4	555	16	Q88L17
28	415	17.3	555	16	O9A5Y0
29	413	17.3	439	16	Q9KFA4
30	410.5	17.1	678	16	Q8P6P5
31	409.5	17.1	632	16	Q9HW93
32	408	17.0	499	16	Q8UGL0
33	405.5	16.9	579	16	Q9K632
34	405.5	16.9	632	2	O32440
35	405	16.9	637	16	Q8D3W6
36	403	16.8	539	16	Q87QG4
37	403	16.8	678	16	Q8PHZ9
38	401	16.8	530	16	Q9XOM7
39	401	16.8	629	16	Q9HW91
40	400.5	16.7	553	16	Q9KNF6
41	400	16.7	664	16	Q8RBV6
42	399.5	16.7	535	16	Q91OR3
43	399.5	16.7	563	16	Q81XF7
44	399	16.7	702	16	Q97D01
45	398	16.6	623	16	Q8D3R3

ALIGNMENTS

RESULT 1

P71412					
ID	P71412	PRELIMINARY;	PRT;	482	AA.
AC	P71412;				
DT	01-FEB-1997 (TREMBLrel. 02, Created)				
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)				
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
DE	Transducer Hta protein.				
GN	HTA				
OS	Halobacterium salinarum.				
OC	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;				
OC	Halobacteriaceae; Halobacterium.				
OX	NCBI_TaxID=2242;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=FLX15;				
RX	MEDLINE=96209786; PubMed=8643458;				
RA	Zhang W., Brooun A., McCandless J., Banda P., Alam M.;				
RT	"Signal transduction in the archaeon Halobacterium salinarum is processed through three subfamilies of 13 soluble and membrane-bound transducer proteins.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).				
DR	EMBL; U75435; AABI7880.1; -				
DR	PIR; T44973; T44973.				
DR	HSSP; P02942; 1QU7.				
DR	GO; GO:0016020; C:membrane; IEA.				
DR	GO; GO:0004871; F:signal transducer activity; IEA.				
DR	GO; GO:0005935; P:chemotaxis; IEA.				
DR	GO; GO:0007165; P:signal transduction; IEA.				
DR	InterPro; IPR004089; Chmtaxis_transd.				
DR	InterPro; IPR003660; HAMP.				
DR	InterPro; IPR004090; Me chemotaxis.				
DR	InterPro; IPR000014; PAS domain.				
DR	Pfam; PF00015; MCPsignal; 1.				
DR	PRINTS; PR00260; CHEMTRNSDUCR.				
DR	SMART; SM00304; HAMP; 1.				
DR	SMART; SM00283; NA; 1.				
DR	SMART; SM00091; PAS; 1.				
DR	PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.				

O59634	halobacteri
Q9UXS0	haloarcula
Q9HS86	halobacteri
O93644	halobacteri
Q9HPA2	halobacteri
Q9AB06	caulobacter
Q9HQT5	halobacteri
O93642	halobacteri
Q815U4	bacillus ce
Q9K669	bacillus ha
Q88L17	pseudomonas
Q9A5Y0	caulobacter
Q9KFA4	bacillus ha
Q8P6P5	xanthomonas
Q9HW93	pseudomonas
Q8UGL0	agrobacteri
Q9K632	bacillus ha
O32440	pseudomonas
Q8D3W6	vibrio vuln
Q87QG4	vibrio para
Q8PHZ9	xanthomonas
Q9XOM7	thermotoga
Q9HW91	pseudomonas
Q9KNF6	vibrio chol
Q8RBV6	thermoanaer
Q91OR3	pseudomonas
Q81XF7	bacillus an
Q97D01	clostridium
Q8D3R3	vibrio vuln

DR PROSITE; PS50112; PAS; 1.
SQ SEQUENCE 482 AA; 50872 MW; ECBEF79B3374CEC7 CRC64;
Query Match 25.5%; Score 609.5; DB 1; Length 482;
Best Local Similarity 37.2%; Pred. No. 1.2e-18;
Matches 181; Conservative 70; Mismatches 197; Indels 39; Gaps 12;
QY 15 NGIDGHALADRGDGLDAEATWLSFTGIDDDTWAALAAEQPLEFATADAL-VTDFVDHLE 73
DB 19 NTLVDPAFA--VDADGAVVAV-----DDQIAAL-----LTAPEDAIGVTDIGERLN 63
QY 74 SYERTQDLFANSPTKTVQQLKETQAEYLLGLGRGEYDTEVAAQARIGKIHDLVLGLGPDVY 133
DB 64 D-----DGSRALANKVADTFIDAHHEY---DGVGLADESALLTGCDVYEDTTVAGNTDLW 116
QY 134 LGATRYTYTGLLDALADDVVVADRGEEAAAADVFLVFLPMLKIL-----TFDQOI--- 184
DB 117 FIATPVYHTGEFRGVI-EIVQDR-SSSARYQSBLQALFGLVDTLDAYDAGRFDAVDIA 174
QY 185 AMDTYIDSVAQRHLHDSRQELANAVATHEAPLSLEATSDVAERTDTMREARTDDOV 244
DB 175 AEDTLIDDEYIQGRNLTFGDTLAAHITVHNVERLEAASQAVSSESAIDELSTAQS 234
QY 245 DRMADYSRISVSASVEEVASTADDVVRTSETDAEALAQOGEAAADDALATMTDIDEATD 304
DB 235 TNVSTVATVETLSATVQEIASTADEVDTSATAERLADDDGSAASDAADMMADVATAAD 294
QY 305 GVTAGVEQIGERAADVSTGVVIDDIAEQTNLMALNASIEAAAGBAGGFAVVADEVKA 364
DB 295 SVTSDVEALQNRLEDIDEVDVITGIAEQTNLMALNASIEAAPGEGEGFAVVADEVKA 354
QY 365 LAESREQSTRVEELVEQAEETVDQDLDEVNQIGEGVERVEEAMETLQBITDAVE- 423
DB 355 LAEDAQSNAGHTBSLSEIQRDADTVTLVTITDRIEDAVAGVEDAMASFEIIVTAVEA 414
QY 424 --DAAGSMQEVSTATDEQAVSTVEVAMVDGVDDRAGEIAAALDDIADATDQOVRTVEEV 481
DB 415 TAERASNM--VSDATNEQASAEIEAAMWDETADLADDTTAVADIQVSTQEAQSMHLDL 472
QY 482 RETVGKL 488
DB 473 DESVSEL 479
RESULT 2
Q9HP10 PRELIMINARY; PRT; 633 AA.
AC Q9HP10;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hcr3 transducer.
GN HTR3 OR VNG1856G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leitauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.P., Pohlschroder M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
RT "Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005085; AAC20060.1; -;
DR PIR; H84336; H84336.
DR HSSP; P02942; 1QU7.

DR GO; GO:0016020; C-membrane; IEA.
DR GO; GO:0004871; F-signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR003660; HAMF.
DR InterPro; IPR004090; Me_chemotaxis.
DR Pfam; PF00672; HAMF; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMF; 2.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSUC_2; 1.
DR PROSITE; PS50885; HAMF; 2.
KW Complete proteome.
SQ SEQUENCE 633 AA; 66423 MW; C4D8022B66ECD0FF CRC64;
Query Match 24.9%; Score 597; DB 17; Length 633;
Best Local Similarity 33.8%; Pred. No. 5.7e-18;
Matches 150; Conservative 76; Mismatches 158; Indels 60; Gaps 7;
QY 42 IDDDTWAALAAEQPLEFATADALVTFDYPHLSYERT----QDLFANSPTKTVQQLKETQA 97
DB 213 LDDELPGAFGASLSQWHTREALITDLDEAREDAEQTRKDAEEAARASERLNERLERRAA 272
QY 98 EYLLGLGRGEYDTEVAAQARIGKIHDLVLGLGPDVVLGAVTRYTYTGLLDALADDVVADRG 157
DB 273 -----EYSDMAAAA-----GDLTR-----RLDEDVDSEPM 299
QY 158 EBAAAADELVARFLPMLKLLTFDQIAMDVTIDSYAQRHLHDSRQELANAVATHVEA 217
DB 300 QDIAEAFNDWMG-----DVEATL-----AQVRSIADAVD-----AASDVTST 336
QY 218 PLSLEATSDVAERTDTMREARTDQVDRMADVSRISVSASVEEVASTADDVVRTSED 277
DB 337 SAAETRSASDQVSESQDISADADQQRDLGTGDEVTSLSATVEEIAASADVAETVQ 396
QY 278 AEALAQOGEAAADDALATMTDIDEATDGTAGVEQLGERAADVESVTGVIDDDIAEQTNML 337
DB 397 ATESERGQELGEDAVAELERIEATDAVERVTALEEAVIDGVTGVTIDTAEQNML 456
QY 338 ALNASIEAARAGEAGGFAVVADEVKALAESESEQSTRVEELVEQMAETETVDQIDEV 397
DB 457 ALNANIEAARADKSGDGFVAVVADEVKDLADEVKESATEIETLVDDVQADVADTADMSEL 516
QY 398 NORIGEGVERVEEAMETLQBITDAVEDAASGMQEVSTATDEQAVSTEEVEAMVDGVDDRA 457
DB 517 GDRVDAGSETIEAALAAALDDIGDQVEAANGSVQISIDATDEQAASTEVEVVTMIDEVTDLS 576
QY 458 GEIAAALDDIADATDQOVRTVEEV 481
DB 577 DRTATESQQVSAAEAEQAASVSEV 600
RESULT 3
Q9P9J0 PRELIMINARY; PRT; 805 AA.
AC Q9P9J0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Chemotaxis transducer protein Bast.
GN BAST.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=59;
RX MEDLINE=20138366; PubMed=10672186;
RX Kokoeva M.V., Oesterhelt D.;
RT "Bast", a membrane-bound transducer protein for amino acid detection in

[illegible]

RESULT 5

```
O93775
ID O93775 PRELIMINARY; PRT; 452 AA.
AC O93775;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Car protein.
GN CAR.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=S9;
RX MEDLINE=99164082; PubMed=10064592;
RA Storch K.F., Rudolph J., Oesterhelt D.;
RT "Car: A cytoplasmic sensor responsible for arginine chemotaxis in the
RT archaeon Halobacterium salinarum.";
RL EMBO J. 18:1146-1158(1999).
DR EMBL; AJ132321; CAB39318.1; -.
DR PIR; T44849; T44849.
DR HSSP; P02942; 1QU7.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR004090; Me chemotaxis.
DR InterPro; IPR000014; PAS domain.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00283; MA; 1.
DR SMART; SM00091; PAS; 1.
DR TIGRFAMs; TIGR00229; sensory box; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
SQ SEQUENCE 452 AA; 49098 MW; 9244D90D694681A CRC64;

Query Match 23.4%; Score 560.5; DB 1; Length 452;
Best Local Similarity 31.3%; Pred. No. 1.4e-16;
Matches 159; Conservative 88; Mismatches 170; Indels 91; Gaps 13;

QY 10 TADVRNGIDGHALADRI-----GLDEAEIA-----WRLGFTGIDDDTMAALAA 52
DB 5 SSDMGGEATGEHLADELCEAYLGDNEDDGGDELQRLSRERDFKHNFN-----QLVA 56
QY 53 RQP--LFEATADALVTD---FYDHLESY-----ERTQDLFANSTKTVQELKETAQYL 100
DB 57 EYPEGILITAAAGCTVTHWNERFSDHMKMARSDALGEDASDVFS---TABE-SETLPEAV 111
QY 101 LGLGRGEYDEYAAQARIGIKIHVLGLGPDVYLGYTRYTGLLDALADDVADRGEAA 160
DB 112 VRTGDTVEEEE-----PHDVPTDSLCOYHGVPLRAPTG-----DVVGSFG--V 152
QY 161 AAADVELVARFLPMLKLLTFDQOIAMDTYIDSYAQRHLHDEIDSRQELANAVATHVEAPLS 220
DB 153 VPDISEKVN-----QRELHD-----LHETVSSNVGEHLS 182
QY 221 SLEATSDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTDVDRTSDEAA 280
DB 183 ELSIESIDVGSFABETEAFAGEIERMEGVADSVNSQATIEBIASSAEVVSQASRAQD 242
QY 281 LAQCEAAADDALATMTDIDEATDGTAGVEQLGERAADVESVTGVDDIDIAEQTNMLAN 340
DB 243 RATEGQTAFATIDRMGAVQESAEVNDITDGLTSQADEMSEIIDAINDIADQTNMLAN 302
QY 341 ASTEERARAGAGGFAVADVEKALAEESREOSQTRVEELVQEQAEETETVDQDEVNQR 400
DB 303 ASLEERARAGEKGGFAVADVEKSLAESQERADEIEQMVENVETTDQADRIGQTTIE 362
QY 401 IGGVSRVEAMETLQBITDAVEDAAGMQEVSTATDEQAQVSTEEVAEMVDGVDDRAGEI 460
DB 363 IEEAITAVRETLDLSQIRKAVDETATGTVKEVAGATDDHAASTEQVAAATTDEAVDKLTEL 422
```

```
QY 461 AAALDDIADATDQOVRTVEEVRETVGKL 488
DB 423 EDRLLNLSQIASSEQHQRVAEIEDMVDEL 450

RESULT 6
P71415 PRELIMINARY; PRT; 777 AA.
ID P71415;
AC P71415;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Transducer Htd protein.
GN HTD.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Fix15;
RX MEDLINE=96209786; PubMed=8643458;
RA Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
RT "Signal transduction in the archaeon Halobacterium salinarum is
RT processed through three subfamilies of 13 soluble and membrane-bound
RT transducer proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).
DR EMBL; U75438; AAB17883.1; -.
DR PIR; T44597; T44597.
DR HSSP; P02942; 1QU7.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me chemotaxis.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 1.
SQ SEQUENCE 777 AA; 82097 MW; 0B20A6A593AA816 CRC64;

Query Match 22.6%; Score 540.5; DB 1; Length 777;
Best Local Similarity 32.4%; Pred. No. 1.8e-15;
Matches 141; Conservative 74; Mismatches 171; Indels 49; Gaps 6;

QY 88 TVEQLKETQAEVLLGLGRGEYDEYAAQRA-RIGKIHVLGLGPDVYLGYTRYTGLLD 146
DB 320 TVRALNDLEAK-APALERGEYDTLDVARVDELGRLEFAFASLRDAVQRADSDANQQVD 378
QY 147 ALA-----DDVVADRGEEAAAADVELVARFLP 173
DB 379 AEARESEAAQAEPPEAAQAEEAEESDAGAPGDDRRGVLGDDACVAGDLTVR--- 435
QY 174 MLKLLTFDQOIA-----MDTYIDSYAQRHLHDEIDSRQELANAVA---THVEAPLSLEAT 226
DB 436 ---LDADVEQAAMADMRRAFNEMADMEATIAERGAFADEVATASTDASDSAAAVEQTG 491
QY 227 QDVAERTDMRTDQVDRMADVSREISSVSASVEEVASTDVDRTSDEAALAQOGE 286
DB 492 RIVSTAVGIRDRRAADQRLQLEAVASETDEMSATIEEVARVAGQVAETSQRAALGDDGQ 551
QY 287 AAADDALATMTDIDEATDGTAGVEQLGERAADVESVTGVDDIDIAEQTNMLANASTEEA 346
DB 552 AAAQDAVAQLELEDETQAAATAVDLEAKMSEIETIVAATIDIAEQTNMLANASTEEA 611
QY 347 RAGEAGEGFAVADVEKALAEESREOSQTRVEELVQEQAEETETVDQDEVNQRIGEGVE 406
```

```

Db 612 RADQDGGFAVAVDEKDLADESKALAAEMQALVAEVRAQTETSVAAAMDRIOERVSQDGE 671
QY 407 RVEEMETLQEIITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVGVDDPAGIAAALDD 466
Db 672 TVSETERSLDIAGRTAEPDGTQVEISNAMDDQRAASVSDVTTAVGDVAALGEETATEAES 731
QY 467 IADATDQQVTVVEEV 481
Db 732 TRDAAAEQATTLSDV 746

RESULT 7
P71414
ID P71414 PRELIMINARY; PRT; 792 AA.
AC P71414;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Transducer Htc protein.
GN HTC.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Flx15;
RA Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
RT "Signal transduction in the archaeon Halobacterium salinarum is
RT processed through three subfamilies of 13 soluble and membrane-bound
RT transducer proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654 (1996).
DR EMBL; U75437; AAB17882.1; -.
DR HSSP; P02942; 1QU7.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_chemotaxis.
DR Pfam; PF00672; HAMP; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; NA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 2.
SQ SEQUENCE 792 AA; 83664 MW; 745D3693F3EEF9F9 CRC64;

Query Match
Best Local Similarity 22.5%; Score 539; DB 1; Length 792;
Matches 143; Conservative 76; Mismatches 163; Indels 62; Gaps 7;

QY 42 IDDDTMAALAAEQPLFEATADALVTDVDFDHLSEYRT-----ODLFANSTKTVQELKETQA 97
Db 374 LDEDLPAGFAGSLSQMH-RLEALITDLDDEARDAEQTRKDAEAAERLNERLERRAA 432
QY 98 EYLLGLGRGEYDTEVAQARIGKIHVLGPDVYLGYTRYTGLLDALADDVVDAG 157
Db 433 -----EYSDENAAAA-----GDLTR-----RLDEDVDSGM 459
QY 158 EEAARAVDELVARFPLMKLLITFDQOIAMDITVDSYAQRLHDEISROELANAVATHVEA 217
Db 460 QDIAEAFNDMGDV-----EATLAQVRSIADAVDAATDVHDAABIR----- 501
QY 218 PLSSLEATSQDVAERTDTMTARTDDQVDRMADVRSISVSASVEVASTADVPTSED 277
Db 502 -----SASDQVESVQDISADAEEQRDLGLTGVGDEVALSATVEDIARPTSPSTVQ 555
QY 278 AEALAQCGEAAADALATMTDIDEATDGTAGVEQLGERAADVESVTGVDDIDABEQTNML 337

```

```

Db 556 AATESRGQELGEDAVAELERIEATADSARVETALBEAIDAIGHVTGVTIDTAEQTNML 615
QY 338 ALMASTEAAPAGAGGFAVAVDEKALAESEREQSTRVEELVEQMAETEETVDQLDEV 397
Db 616 ALNANIEAARADKSGDGFVAVDEKDLADEVKESATEIFLTVDDYQADVADIVADNSEL 675
QY 398 NORIGEGSERVEEAMETLQEIITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVGVDDRA 457
Db 676 GDRVDAGSETIEAALAAVDDIGDQVEAANGSVQISDATDEQAASTEEVVMTIDEVTDLS 735
QY 458 GEIAAALDDIADATDQQVTVVEEV 481
Db 736 DRTATESQQVSARAEQQAASVSEV 759

RESULT 8
Q9HX8
ID Q9HX8 PRELIMINARY; PRT; 636 AA.
AC Q9HX8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Htr15 transducer.
GN HTR15 OR VNG0958G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
DR EMBL; AE005032; AAG19381.1; -.
DR PIR; A84252; A84252.
DR HSSP; P02942; 1QU7.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR004090; Me_chemotaxis.
DR InterPro; IPR000700; PAS-assoc C.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00015; MCPsignal; 1.
DR Pfam; PF00989; PAS; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00283; NA; 1.
DR SMART; SM00091; PAS; 1.
DR TIGRFAMs; TIGR00229; sensory_box; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50113; PAC; 1.
DR PROSITE; PS50112; PAS; 1.
KW Complete proteome.
SQ SEQUENCE 636 AA; 67348 MW; 98D2E415295D4D14 CRC64;

Query Match
Best Local Similarity 22.4%; Score 535.5; DB 17; Length 636;
Matches 149; Conservative 99; Mismatches 191; Indels 63; Gaps 11;

QY 17 IDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT-----ADALVTD----- 67
Db 165 VDDALLDSIGMP-----VFMLDATG-----SVVANNETIELTGCSKADAMGMDHASEA 214
QY 68 FYDHLSEYRTQDLFANSTKTVQELKETQAEYLLGLGRGEYDTEVAQARIGKIHVLG 127

```

Db 215 FPPDRVRKTLAD-----KVIESPRSA-----EPDIERSEEKAQLYADTSVM- 258
 QY 128 LGPDVYLGAITYRYTGLDALADD-----VVADGEEA--AAAVDELVARFLPML- 175
 Db 259 ----TDGGTDRHIRFKASPIFDDDGELLAVAETIEDRTEDVRRADAVEELDELSTIID 314
 QY 176 -----KLLTFDQIAMDITDYSYAQELHDEIDSROELANAVATHVEAPLSLEATSQ 227
 Db 315 ALLSSQGLSKRASFEHIGIINEQLVSVSALNGMADQFERLVGVGDGQTOELADTTIERATA 374
 QY 228 DVAERTDTRARTDQOVDMADVRSREISSVSASVEEASTADDVRRTSDEBALAQOGEA 287
 Db 375 DATDIDATV-----SSQNEMLSSAANEMENFASMQEVAASSQVSAASQAQDAAESGLE 430
 QY 288 AADDALATMTDIDEATDGTAGVQELGERAAVESVTGVDDIIDEAGTNMLALNASIEAAR 347
 Db 431 ASEGANQATNEVIDISDDLMESVSKLESMDIEDVVEIAEVADQTNLLALNANIEAAR 490
 QY 348 AGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMAETEETVDQDDEVNORIGEVER 407
 Db 491 AGEAGSGFAVVADEVKELANETREHTERTAGSISDVQQANETVLAVESHSQIHRAGDE 550
 QY 408 VEEAMETLQEIITDAVEDAASQMOEVSTATDEQAVSTERVAEMVGVDDDRAGEITAAALDDI 467
 Db 551 IDDALTALEIATSDVEAATGITEVARANDEQASTVEDVIVITLEDVQQQAEAEAAASDRI 610
 QY 468 ADATQOQVTRVEEVRTVGKLS 489
 Db 611 VSAEQEQSTAVSGLSERVDKLT 632

RESULT 9

P71416
 ID P71416 PRELIMINARY; PRT; 804 AA.
 AC P71416;
 DT 01-FEB-1997 (T-EMBLrel. 02, Created)
 DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE Transducer Htf protein.
 GN HTF.
 OS Halobacterium salinarum.
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OC NCBI_TaxID=2242;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FLX15;
 RX MEDLINE=96209786; PubMed=8643459;
 RA Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
 RT "Signal transduction in the archaeon Halobacterium salinarum is
 processed through three subfamilies of 13 soluble and membrane-bound
 transducer proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654 (1996).
 DR EMBL; U75439; AAB17884.1; -;
 DR PIR; T44606; T44606.
 DR GO; GO:0016020; C-membrane; IEA.
 DR GO; GO:0004871; F:signal transducer activity; IEA.
 DR GO; GO:0006935; P:chemotaxis; IEA.
 DR GO; GO:0007165; P:signal transduction; IEA.
 DR InterPro; IPR004089; Chmtaxis_trans.
 DR InterPro; IPR003660; HAMP.
 DR Pfam; PF00672; HAMP; 1.
 DR Pfam; PF00015; MCPsignal; 1.
 DR SMART; SM00304; HAMP; 2.
 DR SMART; SM00283; MA; 1.
 DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
 DR PROSITE; PS50885; HAMP; 1.
 SQ SEQUENCE 804 AA; 84705 MW; 0D57284683BFD18B CRC64;

Query Match 22.1%; Score 529; DB 1; Length 804;
 Best Local Similarity 30.6%; Pred. No. 5.9e-15;
 Matches 139; Conservative 76; Mismatches 159; Indels 80; Gaps 9;

QY 89 VEOLKETQAEYLLGLGRGEYTEVAAQR-----ARIGKIHVGLGPDVYLGAITYRY 140
 Db 351 VQSLSAARAE-----LEAGNYDVAATSRDEITGQLFASIGSMRDALVQAE--AAREQA 403
 QY 141 YTGLLDALAD-----DVADRGEEAAAADVELVARFLPMLKLLTFDQOIAMDITDYSIAQR 196
 Db 404 TEAQQDAEADARRAEDARAEDAKADAEALAAE-----LEAQAE 444
 QY 197 LHDEI-----DSR-----OELANAVAT---HV 215
 Db 445 YSDVTGGVADGLTRMPADDTDNEMAAIAASFNDOSWEHTIIDIQEPADAVATASEEA 504
 QY 216 EAPLSLSLATSODVAERTDITMEARTDDQVDRMADVRSREISSVSASVEEASTADDVRRTS 275
 Db 505 EVGAADAERASQVSESQEIAGADEQRNMLDVTSGEMTDLSSAALEEVAS-ADSAEHS 563
 QY 276 EDABALAQOGEAAADALATMTDIDEATDGTAGVQELGERAAVESVTGVDDIIDEAGTN 335
 Db 564 HQTAETIARDGEQTAEDAIERSLTVQEAIDATVQNVLEALDDQMAEISEIVDLISDIAEQTN 623
 QY 336 MLALNASIEAARAGEGFAVVADEVKALAEESREQSTRVEELVEQMAETEETVDQD 395
 Db 624 MLALNANIEAARADKSGGFAVVADEVKDFABEETQESAGDIERRITEVOSQTATVAER 683
 QY 396 EYNORIGEVERVEEAMETLQEIITDAVEDAASQMOEVSTATDEQAVSTERVAEMVGVDD 455
 Db 684 AAESMDAGIDAVEEVVDAFTAVSDHSDETDTGVQEIISDTTDDQAASTEAEVSMTEEVAD 743
 Db 744 LSDSTAGRAQSVSATAEQAASMSISDTVESLS 777

RESULT 10

O06022
 ID O06022 PRELIMINARY; PRT; 451 AA.
 AC O06022;
 DT 01-JUL-1997 (T-EMBLrel. 04, Created)
 DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE Soluble transducer protein Hth.
 DE Halobacterium salinarum.
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OC NCBI_TaxID=2242;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FLX15;
 RX MEDLINE=97284501; PubMed=9139915;
 RA Brooun A., Zhang W., Alam M.;
 RT "Primary structure and functional analysis of the soluble transducer
 protein HtrXI in the Archaeon Halobacterium salinarum.";
 RL J. Bacteriol. 179:2963-2968 (1997).
 DR EMBL; U74668; AAC45264.1; -;
 DR PIR; T44964; T44964.
 DR HSP; P02942; 1QU7.
 DR GO; GO:0016020; C-membrane; IEA.
 DR GO; GO:0004871; F:signal transducer activity; IEA.
 DR GO; GO:0006935; P:chemotaxis; IEA.
 DR GO; GO:0007165; P:signal transduction; IEA.
 DR InterPro; IPR004089; Chmtaxis_trans.
 DR InterPro; IPR004090; Me_chemotaxis.
 DR InterPro; IPR000014; PAS_domain.
 DR Pfam; PF00015; MCPsignal_1.
 DR PRINTS; PR00260; CHEMTRNSDUCR.
 DR SMART; SM00283; MA; 1.
 DR SMART; SM00091; PAS; 1.
 DR TIGRFAMs; TIGR00229; sensory_box; 1.
 DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
 SQ SEQUENCE 451 AA; 49122 MW; 6DA9E4FF8139DA87 CRC64;

Query Match

21.9%; Score 524; DB 1; Length 451;

Best Local Similarity 30.9%; Pred. No. 4.9e-15;		Matches 158; Conservative 87; Mismatches 168; Indels 98; Gaps 16;	
QY	10 TADVRNGIDGHALADRI-----GLDEAEIA-----WRLSFTGIDDDTMAALAA 52	DR	PRINTS; PR00260; CHEMTRNSDUCR.
Db	5 SSDMGATGEHLADELCEAYLGDNDGDELQRLSRERDFWKHMFN-----QLVA 56	DR	SMART; SM00304; HAMP; 1.
QY	53 EOP--LFEATADALVTD----FYDHLESY-----ERTODLFANSTKTVEQLKETQAEYL 100	DR	SMART; SM00283; MA; 1.
Db	57 EYPEGILITADGTVTHWERSFSDMKMARSALGEDASDVFS-----TAE-SETLPEAV 111	DR	PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
QY	101 LGLGRGEYDEYAAQARARIKIHVDVLGPDVYLGAYTRYTYTGLLDALADDVVADRGEA 160	DR	PROSITE; PS00885; HAMP; 1.
Db	112 VRTGDTVEEEE-----PHDVPTSLCQVHGVELRAPTG-----DVVSGFG--V 152	SQ	SEQUENCE 544 AA; 56922 MW; 663D8D0E43AFCA71 CRC64;
QY	161 AAANDVELVARFLPMLKLLTFDQOIAMDTYDSYAQRLHDEIDSRQELANAVATHVEAPLS 220	Query Match 21.9%; Score 523.5; DB 1; Length 544;	
Db	153 VPDISEKVN-----QRELHD-----LHETVSINVGEHL 182	Best Local Similarity 29.4%; Pred. No. 6.4e-15;	
QY	221 SLEATSQDV---AERTDTMRATDDOVRMADVSREISSVSASVEEVASTDVVRTSED 277	Matches 141; Conservative 93; Mismatches 195; Indels 51; Gaps 10;	
Db	183 ELSSEIDKVGSAETEPAGK---EIERMEGFADKVSQSATIEIASAEVQSQR 239	QY	28 LDEATLWELSTGIDDDTMAALAAEQPLFEATADALVTDF----YDHLESYERQDLFA 83
QY	278 AEALAQOGAAADDALATMTDIDEATDGTAGVQLGERAADVESVTGVIDDIAEQTNML 337	Db	64 LKAGVATOI--FNLATGPLMATTVAVFAGGTATTVAIVEDMEALVEERAQPKQEAEEEA 122
Db	240 AQDRATEGEQTAEATIDRMGAQESAERVNDITDGLTSQADEMSEIIDAINDIAQTNML 299	QY	84 NSTKTVEQLKETQAEYLLGLGRGEYDT-EYAAQARARIKIHVDVLGPDVYLGAYTRYTYT 142
QY	338 ALNASTEARAGEAGEGFVAVADEVKALAEBSREQSTRVEELVEQMQAETEFTVDQLDV 397	Db	123 EAERAREKAQKQAE-----AERQTAEAQSGTGTGTQREIEQLAAD--LESQATEVG 173
Db	300 ALNASTEARAGEKGGFVAVADEVKSLAEBSQERADEIEQMVEMVETTDQTADRIGOT 359	QY	143 GLLDALAD-----DVVADRGE--BAAAVDELVARFLPMLKLLTFDQOIAMDTYDSY 193
QY	398 NORIGEGVERVEEAMETLQEIETDAVEDAASGMQEVSTATDEQAVSTEEVAEMVGDVDR 457	Db	174 ATLEAASDGLTARVDATTDNAEIAEAVATVNDM-----LTTMERTI----- 215
Db	360 TTEIEEATVAVRETLDLSQEIINAVDETATGKVEAGARD-HAASTEQVAAATTDEAVDKL 418	QY	194 AQRLHDEIDSRQELANAVATHVEAPLSLSLEATSQVABERTDTMRATDDOVRMADVSRE 253
QY	458 GEIAAALDDIADATDQQVRIVEEVRTVGKL 488	Db	216 -----DEIQGFSTNTTASREATGGAKEIQVASQTVSESVQEIAGTDDQREQLSVAEE 270
Db	419 TELERLDNLNSQIASQEHQDRVAIEDMWDEL 449	QY	254 ISSVSASVEEVASTDVVRTSEDAAEQGEEAADDALATMTDIDEATDGTAGVEQL 313
RESULT 11		Db	271 MDSYATVEEVAATAQSVADTAADTTDVATACKQTAEDAIDAIDAVQETMQTTVANVDAL 330
P71409	PRELIMINARY; PRT; 544 AA.	QY	314 GERAADVESVTGVIDDIAEQTNMLALNASTIEAARAGEA-----GEGFVAVADVKKALAEBS 369
AC	P71409;	Db	331 EDLTTTEIIDIAELISDIAEQTNMLALNASTIEAARAGSGGGTNGDGFVAVVADEVKELATES 390
DT	01-FEB-1997 (TremBLrel. 02, Created)	QY	370 REQSTRVEELVQMOAETEFTVDQLDDEVNORTIGEGVERVEEAMETLQEIETDAVEDAASGM 429
DT	01-FEB-1997 (TremBLrel. 02, Last sequence update)	Db	391 QBSAKDIAELIEVQSQTATTVEEIRVAEQRVNDGAAVAEETVDFAGFATENIQETTDGV 450
DT	01-JUN-2003 (TremBLrel. 24, Last annotation update)	QY	430 QEVSTATDEQAVSTEEVAEMVGDVDRAGEIAAALDDIADATDQQVRIVEEVRTVGKLS 489
DE	Htr8 transducer.	Db	451 QEISQAMDEQAQRSEVSSVDDIATISQATADRAENVSAASEEQTASITETVGSLSQSLA 510
GN	HTR8 OR VNG1523G.	RESULT 12	
OS	Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).	Q9HPQ5	PRELIMINARY; PRT; 643 AA.
OC	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;	AC	Q9HPQ5;
OC	Halobacteriaceae; Halobacterium.	DT	01-MAR-2001 (TremBLrel. 16, Created)
OX	NCBI_TaxID=64091;	DT	01-MAR-2001 (TremBLrel. 16, Last sequence update)
RN	[1]	DT	01-JUN-2003 (TremBLrel. 24, Last annotation update)
RP	SEQUENCE FROM N.A.	DE	Htr8 transducer.
RX	MEDLINE=20504483; PubMed=11016950;	OS	Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
RA	Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,	OC	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
RA	Shukla H.D., Lasky S.R., Baliga N.S., Thorsen V., Sbrogna J.,	OC	Halobacteriaceae; Halobacterium.
RA	Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,	OX	NCBI_TaxID=64091;
RA	Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,	RN	[1]
RA	Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,	RP	SEQUENCE FROM N.A.
RA	Isenbarger T.A., Peck R.F., Pchlschroder M., Spudich J.L., Jung K.-H.,	RX	MEDLINE=20504483; PubMed=11016950;
RA	Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,	RA	Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA	Ebbardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;	RA	Shukla H.D., Lasky S.R., Baliga N.S., Thorsen V., Sbrogna J.,
RT	"Genome sequence of Halobacterium species NRC-1."	RA	Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RL	EMBL; AE005065; AAG19812.1; --	RA	Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
DR	PIR; H84305; H84305.	RA	Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
DR	HSSP; P02942; 1Q07.	RA	Isenbarger T.A., Peck R.F., Pchlschroder M., Spudich J.L., Jung K.-H.,

DR	GO:	GO:0016020;	C:membrane; IEA.		
DR	GO:	GO:0004871;	F:signal transducer activity; IEA.		
DR	GO:	GO:0006935;	P:chemotaxis; IEA.		
DR	GO:	GO:0007165;	P:signal transduction; IEA.		
DR	InterPro:	IPR004089;	Chmtaxis_transd.		
DR	InterPro:	IPR003660;	HAMP.		
DR	InterPro:	IPR004090;	Me_chemotaxis.		
DR	Pfam:	PF00015;	MCPsignal; 1.		
DR	PRINTS:	PR00260;	CHEMTRNSDUCR.		
DR	SMART:	SMO0304;	HAMP; 1.		
DR	SMART:	SMO0283;	MA; 1.		
DR	PROSITE:	PS50111;	CHEMOTAXIS_TRANSDUC_2; 1.		
DR	PROSITE:	PS50885;	HAMP; 1.		
KW	Complete proteome.				
SQ	SEQUENCE	643 AA;	67290 MW; 1D5F2E1BBA02481 CRC64;		
Query Match 21.9%; Score 523.5; DB 17; Length 643;					
Best Local Similarity 28.3%; Pred. No. 7.8e-15;					
Matches 143; Conservative 92; Mismatches 200; Indels 71; Gaps					
Qy	8	LVTADVNRGIDGHALADRIGLDEAEFIARWLSTFGTIDDTTMAALAABQLFEATADALVTD	67		
Db	156	LVYVVFTGVGFMINAEVRYNHTAAINPNPWGGIHGAFLVLLAG-----ALMAN	205		
Qy	68	FYDHLESYERTQDLFANSTKTKVEQLKQAELYLLGLGRGEYTEYA-----	113		
Db	206	WYSTERSPEAQSKLREARQAQQVEDLEAR-----QAEIEAKAEAKRLKADAEEARE	259		
Qy	114	---AQRARIGKIHDVLGPDVYLGYTRYTYGLLDALAD-DVVDARGEEAAAADVELVA	169		
Db	260	AAEAQQREVAAALNLEATANTYGAAMARAADDGLSVRLDPDVENDAMAALIAASFNEML-	318		
Qy	170	RFLPMLKLLTFDQQIAMDITYDSYAQLRHDEIDSRLQELANAVATHV-----EAPLSL	222		
Db	319	-----DETETIRETOAVSDVAASESDADAGVVEI	349		
Qy	223	EATSQDVAERTDTMPARTDDQDMRADYSRSISSVASVEEVASTADDVRRTSDEDAALA	282		
Db	350	EDASQVSSETVQETIAAGADEVQEKLETSGEMTDLSSATEEVAAGSADSVAERSHETAAVA	409		
Qy	283	QOGENAADDALATMTDIDEATDGTVTAGVEQLGERAADVESVTGVDDIDIAEQTNLMALNAS	342		
Db	410	GGEGQTATQADSTVSQSVESTQNVEALDDQLAEISEIVDLISDVAEQTNLMALNAN	469		
Qy	343	TEAARAGAGEGFVAVADEVKALABESREQSRVBELVEMQOAETEETVDQLDEVNQRIC	402		
Db	470	TEAARADKSGDFVAVADEVKDLAEETRASAGDIEALVADIDAQMQATVTEARTADESVQ	529		
Qy	403	EGVERVEAMETLQEITDAVEDAAGSMQBVSSTATDUEQAVSTEVAWMVGCVDDRAGEIAA	462		
Db	530	DAISAVIDAVDPAGFVNAENAEETDTGVQISITTTDDQAASTEAVASMIAEVSDISTATAA	589		
Qy	463	ALLDDIADATDQOVRTVEEVRETGVKL	488		
Db	590	DAQQAATAEQQTAAATISENTAAL	615		
RESULT 13					
Q9HRN6 PRELIMINARY; PRT; 628 AA.					
ID	Q9HRN6				
AC	Q9HRN6;				
DT	01-MAR-2001	(TREMBLrel. 16, Created)			
DT	01-JUN-2001	(TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)			
DE	Htr16 transducer.				
GN	HTR16 OR VNG0614G.				
OS	Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).				
OC	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;				
OC	Halobacteriaceae; Halobacterium.				
OX	NCBI_TaxId=64091;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=20504483; PubMed=11016950;				

QY 439 QAVSTEEVAEMVDCYDDRAGEIAAALDDIADATDQOVRTVEEV 481
DB 702 QADATQSVVRRYDDVADISQHVTEDAEQVSAAAEEQASVAEI 744

Search completed: August 10, 2004, 15:28:39
Job time : 60.609 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:18:30 ; Search time 26.4879 Seconds
(without alignments)
1632.058 Million cell updates/sec

Title: US-09-455-978B-76

Perfect score: 794

Sequence: 1 VLSEGEWQLVHWAKVEAD.....ALELPKDIATAKYKELGYQG 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	794	100.0	153	2 AAY29391	Aay29391 Sperm wha
2	794	100.0	153	2 AAW81769	Aaw81769 Whale MYG
3	794	100.0	153	3 AAY69975	Aay69975 MYGL prot
4	715	90.1	154	2 AAW62270	Aaw62270 Modified
5	713	89.8	154	2 AAW29741	Aaw29741 Modified
6	711	89.5	154	2 AAW29739	Aaw29739 Modified
7	711	89.5	154	2 AAW29740	Aaw29740 Modified
8	711	89.5	154	2 AAW62271	Aaw62271 Modified
9	707	89.0	154	2 AAW29742	Aaw29742 Modified
10	701	88.3	154	2 AAW29743	Aaw29743 Modified
11	700	88.2	154	2 AAW62272	Aaw62272 Modified
12	699	88.0	154	2 AAW29744	Aaw29744 Modified
13	699	88.0	166	7 ADC31106	Adc31106 Human nov
14	699	88.0	206	7 ADC32817	Adc32817 Human nov
15	697	87.8	154	2 AAW62273	Aaw62273 Modified
16	613	77.2	202	4 ABG04708	Abg04708 Novel hum
17	535	67.4	231	4 ABG04706	Abg04706 Novel hum
18	472	59.4	113	3 AAG00022	Aag00022 Human sec
19	386.5	48.7	334	4 ABG21418	Abg21418 Novel hum
20	314	39.5	71	3 AAG00021	Aag00021 Human sec
21	253	31.9	107	4 ABG21417	Abg21417 Novel hum
22	243	30.6	64	2 AAR95437	Aar95437 Apomyoglo
23	240	30.2	60	2 AAR95436	Aar95436 Apomyoglo
24	215.5	27.1	190	5 ABB07967	Abb07967 Rat STUF
25	215.5	27.1	201	2 AAW86334	Aaw86334 Kidney in

ALIGNMENTS

RESULT 1

AAY29391
ID AAY29391 standard; protein; 153 AA.

XX AC AAY29391;

XX DT 01-OCT-1999 (first entry)

XX DE Sperm whale myoglobin protein sequence.

XX KW Myoglobin; Mb; sperm whale; protein design; function; property;
XX KW stereo structure; globular protein; detection.

XX OS Physeter sp.

XX PN JP11193297-A.

XX PD 21-JUL-1999.

XX PF 06-OCT-1998; 98JP-00283852.

XX PR 06-OCT-1997; 97JP-00272431.

XX PA (RIKA) RIKAGAKU KENKYUSHO.

XX DR WPI; 1999-462430/39.

XX PT Method for designing artificial protein - useful for producing proteins with required functions.

XX PS Example; Page 4-5; 10pp; Japanese.

XX CC The present invention describes a method for producing an ideal amino acid sequence. The method comprises: (1) preparation of one initial amino acid sequence corresponding to the structure of a protein; (2) selecting the amino acid residues optimal to each site; (3) repeating the steps of selecting the amino acid residues optimum to each site of the total amino acid residues constituting the N-order amino acid sequence (N is an integer not less than 2) to give an N+1-order amino acid sequence consisting of the selected amino acid residues until the N-order amino acid sequence comes to be same as the N+1-order amino acid sequence; and (4) selecting the resultant N-order amino acid sequence as the optimum amino acid sequence of said protein. The method can design a protein with desired functions and properties. The present sequence represents sperm whale myoglobin used in an example from the present invention where the stereo structure of sperm whale myoglobin was targeted as a globular protein to detect an amino acid sequence best fit to the structure of the main chain of the sperm whale myoglobin

Aae04661 Haem prot
Aar95435 Apomyoglo
Abb07966 Human STU
Adc10192 Human NOV
Adc10194 Human NOV
Aae04662 Haem prot
Aae04665 Haem prot
Aae04664 Haem prot
Aae04669 Haem prot
Aae04670 Haem prot
Aae04673 Haem prot
Aae04672 Haem prot
Aae04663 Haem prot
Aay29393 Sperm wha
Aae04671 Haem prot
Aab41929 Human ORF
Aae04674 Haem prot
Abg04707 Novel hum
Abg04705 Novel hum
Aaw73385 Human hae

```

XX SQ Sequence 153 AA;
Query Match 100.0%; Score 794; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 3e-80;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKMSD 60
Db 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKMSD 60

QY 61 LKXHGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
Db 61 LKXHGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120

QY 121 GDFGADAQAGAMNKALELFKDKIAAKYKELGYQG 153
Db 121 GDFGADAQAGAMNKALELFKDKIAAKYKELGYQG 153

RESULT 2
AAW81769
ID AAW81769 standard; peptide; 153 AA.
XX
AC AAW81769;
XX
XX 29-JAN-1999 (first entry)
XX
DE Whale MYGL peptide fragment.
XX
KW DHFR; dihydrofolic acid reductase; protein function; trypsin; bovine;
KW ribonuclease; myoglobin; database; homology; resemblance; whale.
XX
OS Cetacea.
XX
XX JP10287696-A.
XX
XX 27-OCT-1998.
XX
XX 11-APR-1997; 97JP-00093577.
XX
XX 11-APR-1997; 97JP-00093577.
XX
XX (IYAK-) IYAKU BUNSHI SEKKEI KENKYUSHO KK.
XX
XX WPI; 1999-018384/02.
XX
XX Determination of protein biological function - comprises use of amino
XX acid sequences database containing the relevant information.
XX
XX Example 1; Fig 1; 1lpp; Japanese.
XX
XX This sequence is used in the creation of a database containing the
XX information for amino acid sequence of protein with at least 1 biological
XX function with added a score on importance of expression of the biological
XX information for each amino acid residue. The database is useful for
XX determination of unknown biological function of a protein or polypeptide
XX based on the homology of amino acid sequence, e.g. steric structure of
XX protein, and includes retrieval and evaluation of high homologous
XX relationship for the determination of mostly resembling protein. The
XX database allows for correct and rapid retrieval and presumption of
XX protein and polypeptide having biological functions
XX
XX Sequence 153 AA;
Query Match 100.0%; Score 794; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 3e-80;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKMSD 60
Db 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKMSD 60

```

```

QY 61 LKXHGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
Db 61 LKXHGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120

QY 121 GDFGADAQAGAMNKALELFKDKIAAKYKELGYQG 153
Db 121 GDFGADAQAGAMNKALELFKDKIAAKYKELGYQG 153

RESULT 3
AA69975
ID AA69975 standard; protein; 153 AA.
XX
AC AA69975;
XX
XX 13-APR-2000 (first entry)
XX
DE MYGL protein.
XX
KW Protein sequence database; biological function determination;
KW enzymatic activity; signaling activity; protein function determination;
KW MYGL protein.
XX
XX Unidentified.
XX
XX WO9962004-A1.
XX
XX 02-DEC-1999.
XX
XX 26-MAY-1998; 98WO-JP002302.
XX
XX 26-MAY-1998; 98WO-JP002302.
XX
XX (MEDI-) INST MEDICINAL MOLECULAR DESIGN INC.
XX
XX Itai A, Tomioka N, Itai R, Imamura M;
XX WPI; 2000-136797/12.
XX
XX An efficient, accurate and rapid computer database for estimating protein
XX functions e.g. enzymatic activity, for polypeptides obtained from gene
XX sequence translation.
XX
XX Example 1; Fig 1; 26pp; Japanese.
XX
XX This sequence represents the MYGL protein. The invention relates to a
XX database containing information on the amino acid (aa) sequences of
XX proteins of which 1 or more biological functions are known. The database
XX also contains additional information on the score of importance of each
XX aa residue in the whole aa sequence in determining the known biological
XX functions. The invention also relates to a method of preparing an
XX alignment between aa sequences contained in the database and those of the
XX unknown polypeptide. This is represented as the homology amongst various
XX sites, each being identified as having a high score of importance in
XX determining potential biological functions. The method is used to enable
XX an efficient estimation of the biological functions (particularly
XX enzymatic and signaling activities) of polypeptides from their aa
XX sequences. Suitable proteins can then be isolated and purified by various
XX means. This could be of considerable use in a biological and medical
XX context. The computerised procedure is efficient, fast and accurate
XX
XX Sequence 153 AA;
Query Match 100.0%; Score 794; DB 3; Length 153;
Best Local Similarity 100.0%; Pred. No. 3e-80;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKMSD 60
Db 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKMSD 60

QY 61 LKXHGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120

```

Qy 2 LSEGEWOLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKXDFRKHILKTEAEKMSEDL 61

Db	3	LSDGEWQVNLVNWGKVEADIAHGQEVLRIRLFTGHPETLEKFDKFKHLKTEAEKASDDL	62
QY	62	KKHGTVTLTALGAILKKKGHEAEKPLAQSHATKHPIKYLEFISEAIHVLHSRHPG	121
Db	63	KKHGTVTLTALGAILKKKGHEAEKPLAQSHATKHPIKYLEFISDAIHVLHSKHFG	122
QY	122	DFGADAQGMNVALELFRKDIAAKYKELGYQG	153
Db	123	DFGADAQGMNTKALELFRNDIAAKYKELGFGQ	154
RESULT 6			
AAW29739	standard; protein; 154 AA.		
ID	AAW29739		
XX	AAW29739;		
AC	26-OCT-1998 (first entry)		
DT	Modified myoglobin protein 5.		
XX	Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);		
XX	Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyze;		
KW	phenolic; polymer; antigen; antibody; diagnostic assay.		
KW	Equus sp.		
XX	Synthetic.		
OS	CA2185605-A.		
XX	17-MAR-1998.		
PN	16-SEP-1996; 96CA-02185605.		
XX	16-SEP-1996; 96CA-02185605.		
PD	(MAUK/) MAUK A G.		
XX	(WANL/) WAN L.		
PA	(LEEH/) LEE H.		
PA	(BRAY/) BRAYER G D.		
PA	(TONG/) TONG H.		
PA	(SMIT/) SMITH M.		
XX	Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;		
PI	WPI; 1998-388655/34.		
XX	New modified myoglobin with amino acid alterations has increased		
PT	peroxidase activity - used for catalysing oxidation of substrates with		
PT	peroxides, e.g. in waste treatment and as label for antigens.		
XX	Claim 10; Page; 23pp; English.		
PS	Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic		
XX	horse heart myoglobin (Mb) proteins, produced by random mutagenesis of		
CC	the synthetic Mb gene. This present sequence contains a substitution at		
CC	Thr391le which has been shown to enhance the peroxidase activity of Mb.		
CC	These protein can be used to catalyze oxidation of a substrate by		
CC	peroxidase, e.g. oxidation of Mn2+. Peroxidases are used for oxidising		
CC	phenolics to polymers, as labels for e.g. antigens, antibodies or		
CC	oligonucleotides, in diagnostic assays and for oxidative treatment of		
CC	waste streams. NB. This sequence is not given in the specification but		
CC	was created from the wild type sequence given by the inventors. The		
CC	position of the substitution given does not include the initiation codon		
XX	Sequence 154 AA;		
SQ	Query Match 89.5%; Score 711; DB 2; Length 154;		
	Best Local Similarity 87.5%; Pred. No. 5.5e-71;		
	Matches 133; Conservative 9; Mismatches 10; Indels 0; Gaps 0;		
QY	2	LSEGEWQLVHLVWAKVADVAGHGQDILIRLFKSHPTLEKFDKFKHLKTEAEKASDDL	61

Db 3 LSGEQVQLVNWGVKVEADVAGHGQVILIRLFTGHPETLEKFKDKLHLKTEAEMKASEDL 62
QY 62 KHGVTVLTAAGAILKKKGHEAEALPLAQSHATKHKIPKYLEFISEAIHVLHSHRHPG 121
Db 63 KHGVTVLTAAGAILKKKGHEAEALPLAQSHATKHKIPKYLEFISEAIHVLHSHRHPG 122
QY 122 DFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
Db 123 DFGADAQAGAMTKALELFRNDIAAKYKELGFQG 154

RESULT 8
AAW62271
ID AAW62271 standard; protein; 154 AA.
AC AAW62271;
XX
DT 26-OCT-1998 (first entry)
DE Modified myoglobin protein 2.
DE
KW Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);
KW Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyze;
KW phenolic; polymer; antigen; antibody; diagnostic assay.
XX
OS Equus sp.
OS Synthetic.
XX
PN CA2185605-A.
XX
PD 17-MAR-1998.
XX
PF 16-SEP-1996; 96CA-02185605.
XX
PR 16-SEP-1996; 96CA-02185605.
XX
PA (MAUK/) MAUK A G.
PA (WANL/) WAN L.
PA (LEEH/) LEE H.
PA (BRAY/) BRAYER G D.
PA (TONG/) TONG H.
PA (SMIT/) SMITH M.
XX
PI Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;
XX WPI; 1998-388655/34.
XX
PT New modified myoglobin with amino acid alterations has increased
PT peroxidase activity - used for catalysing oxidation of substrates with
PT peroxides, e.g. in waste treatment and as label for antigens.
XX
PS Claim 6; Page; 23pp; English.
XX
CC Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic
CC horse heart myoglobin (Mb) proteins, produced by random mutagenesis of
CC the synthetic Mb gene. This present sequence contains a substitution at
CC Lys45Glu, which has been found to enhance the peroxidase activity of the
CC protein in general, and it may also exhibit enhanced Mn(II) peroxidase
CC activity as well. This is due to the enhanced rate of reaction of Fe(III)
CC Mb with hydrogen peroxide and also because the substitution at residue 45
CC creates a new Mn binding site which is stronger than the natural Mn
CC binding site of Mb. This protein also contains a substitution at
CC Lys63Glu, which was also found to be required for Mn binding. These
CC protein can be used to catalyze oxidation of a substrate by peroxidase,
CC e.g. oxidation of Mn2+. Peroxidases are used for oxidising phenolics to
CC polymers, as labels for e.g. antigens, antibodies or oligonucleotides, in
CC diagnostic assays and for oxidative treatment of waste streams. NB. This
CC sequence is not given in the specification but was created from the wild
CC type sequence given by the inventors. The positions of the substitutions
CC given do not include the initiation codon

Sequence 154 AA;

Query Match 89.5%; Score 711; DB 2; Length 154;
Best Local Similarity 87.5%; Pred. No. 5.5e-71;
Matches 133; Conservative 9; Mismatches 10; Indels 0; Gaps 0;
QY 2 LSGEQVQLVNWGVKVEADVAGHGQVILIRLFTGHPETLEKFKDKLHLKTEAEMKASEDL 61
Db 3 LSGEQVQLVNWGVKVEADVAGHGQVILIRLFTGHPETLEKFKDKLHLKTEAEMKASEDL 62
QY 62 KHGVTVLTAAGAILKKKGHEAEALPLAQSHATKHKIPKYLEFISEAIHVLHSHRHPG 121
Db 63 KHGVTVLTAAGAILKKKGHEAEALPLAQSHATKHKIPKYLEFISEAIHVLHSHRHPG 122
QY 122 DFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
Db 123 DFGADAQAGAMTKALELFRNDIAAKYKELGFQG 154

RESULT 9
AAW29742
ID AAW29742 standard; protein; 154 AA.
XX
AC AAW29742;
XX
DT 26-OCT-1998 (first entry)
DE Modified myoglobin protein 8.
DE
KW Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);
KW Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyze;
KW phenolic; polymer; antigen; antibody; diagnostic assay.
XX
OS Equus sp.
OS Synthetic.
XX
PN CA2185605-A.
XX
PD 17-MAR-1998.
XX
PF 16-SEP-1996; 96CA-02185605.
XX
PR 16-SEP-1996; 96CA-02185605.
XX
PA (MAUK/) MAUK A G.
PA (WANL/) WAN L.
PA (LEEH/) LEE H.
PA (BRAY/) BRAYER G D.
PA (TONG/) TONG H.
PA (SMIT/) SMITH M.
XX
PI Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;
XX WPI; 1998-388655/34.
XX
PT New modified myoglobin with amino acid alterations has increased
PT peroxidase activity - used for catalysing oxidation of substrates with
PT peroxides, e.g. in waste treatment and as label for antigens.
XX
PS Claim 10; Page; 23pp; English.
XX
CC Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic
CC horse heart myoglobin (Mb) proteins, produced by random mutagenesis of
CC the synthetic Mb gene. This present sequence contains substitutions at
CC Phe46Leu and Ile107Phe which have been shown to enhance the peroxidase
CC activity of Mb. These protein can be used to catalyze oxidation of a
CC substrate by peroxidase, e.g. oxidation of Mn2+. Peroxidases are used for
CC oxidising phenolics to polymers, as labels for e.g. antigens, antibodies
CC or oligonucleotides, in diagnostic assays and for oxidative treatment of
CC waste streams. NB. This sequence is not given in the specification but
CC was created from the wild type sequence given by the inventors. The
CC positions of the substitutions given do not include the initiation codon

Sequence 154 AA;

Query Match 89.0%; Score 707; DB 2; Length 154;
 Best Local Similarity 86.8%; Pred. No. 1.5e-70;
 Matches 132; Conservative 9; Mismatches 11; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDL 61
 DB 3 LSGEWOQVNLVWGKVEADIAHGQEVLRFTGHPETLEKDFKHLKTEAEKASEDL 62

QY 62 KKHGTVTLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIHVLHSHRHPG 121
 DB 63 KKHGTVTLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISDAIHVLHSHRHPG 122

QY 122 DFGADAQAGAMNKALELFRKIDIAAKYKELGYQG 153
 DB 123 DFGADAQAGAMTKALELFRNDIAAKYKELGFQG 154

RESULT 10
 AAW29743
 ID AAW29743 standard; protein; 154 AA.
 XX
 AC AAW29743;
 XX
 DT 26-OCT-1998 (first entry)
 XX
 DE Modified myoglobin protein 9.
 XX
 KW Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);
 XX Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyse;
 KW phenolic; polymer; antigen; antibody; diagnostic assay.
 XX
 OS Equus sp.
 OS Synthetic.
 XX
 PN CA2185605-A.
 XX
 PD 17-MAR-1998.
 XX
 PF 16-SEP-1996; 96CA-02185605.
 XX
 PR 16-SEP-1996; 96CA-02185605.
 XX
 PA (MAUK/) MAUK A G.
 PA (WANL/) WAN L.
 PA (LEEH/) LEE H.
 PA (BRAY/) BRAYER G D.
 PA (TONG/) TONG H.
 PA (SMIT/) SMITH M.
 XX
 PI Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;
 XX
 DR WPI; 1998-388655/34.
 XX
 PT New modified myoglobin with amino acid alterations has increased
 PT peroxidase activity - used for catalysing oxidation of substrates with
 PT peroxides, e.g. in waste treatment and as label for antigens.
 XX
 PS Claim 12; Page; 23pp; English.
 XX
 CC Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic
 CC horse heart myoglobin (Mb) proteins, produced by random mutagenesis of
 CC the synthetic Mb gene. This present sequence contains substitutions at
 CC Thr35Ile, Phe46Leu, and Ile107Phe, which have been shown to enhance the
 CC peroxidase activity of Mb. These protein can be used to catalyse
 CC oxidation of a substrate by peroxidase, e.g. oxidation of Mn2+.
 CC Peroxidases are used for oxidising phenolics to polymers, as labels for
 CC e.g. antigens, antibodies or oligonucleotides, in diagnostic assays and
 CC for oxidative treatment of waste streams. NB. This sequence is not given
 CC in the specification but was created from the wild type sequence given by
 CC the inventors. The positions of the substitutions given do not include
 CC the initiation codon
 XX
 SQ Sequence 154 AA;

Query Match 88.3%; Score 701; DB 2; Length 154;
 Best Local Similarity 86.2%; Pred. No. 7.2e-70;
 Matches 131; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDL 61
 DB 3 LSGEWOQVNLVWGKVEADIAHGQEVLRFTGHPETLEKDFKHLKTEAEKASEDL 62

QY 62 KKHGTVTLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIHVLHSHRHPG 121
 DB 63 KKHGTVTLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISDAIHVLHSHRHPG 122

QY 122 DFGADAQAGAMNKALELFRKIDIAAKYKELGYQG 153
 DB 123 DFGADAQAGAMTKALELFRNDIAAKYKELGFQG 154

RESULT 11
 AAW62272
 ID AAW62272 standard; protein; 154 AA.
 XX
 AC AAW62272;
 XX
 DT 26-OCT-1998 (first entry)
 XX
 DE Modified myoglobin protein 3.
 XX
 KW Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);
 XX Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyse;
 KW phenolic; polymer; antigen; antibody; diagnostic assay.
 XX
 OS Equus sp.
 OS Synthetic.
 XX
 PN CA2185605-A.
 XX
 PD 17-MAR-1998.
 XX
 PF 16-SEP-1996; 96CA-02185605.
 XX
 PR 16-SEP-1996; 96CA-02185605.
 XX
 PA (MAUK/) MAUK A G.
 PA (WANL/) WAN L.
 PA (LEEH/) LEE H.
 PA (BRAY/) BRAYER G D.
 PA (TONG/) TONG H.
 PA (SMIT/) SMITH M.
 XX
 PI Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;
 XX
 DR WPI; 1998-388655/34.
 XX
 PT New modified myoglobin with amino acid alterations has increased
 PT peroxidase activity - used for catalysing oxidation of substrates with
 PT peroxides, e.g. in waste treatment and as label for antigens.
 XX
 PS Claim 23; Page; 23pp; English.
 XX
 CC Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic
 CC horse heart myoglobin (Mb) proteins, produced by random mutagenesis of
 CC the synthetic Mb gene. This present sequence contains a substitution at
 CC Lys45Glu, which has been found to enhance the peroxidase activity of the
 CC protein in general, and it may also exhibit enhanced Mn(II) peroxidase
 CC activity as well. This is due to the enhanced rate of reaction of Fe(III)
 CC Mb with hydrogen peroxide and also because the substitution at residue 45
 CC creates a new Mn binding site which is stronger than the natural Mn
 CC binding site of Mb. This protein also contains a substitution at
 CC Lys63Glu, which was also found to be required for Mn binding, and a
 CC substitution at His97Leu to increase the rate of Mn(II) turnover. These
 CC protein can be used to catalyse oxidation of a substrate by peroxidase,
 CC e.g. oxidation of Mn2+. Peroxidases are used for oxidising phenolics to

CC polymers, as labels for e.g. antigens, antibodies or oligonucleotides, in
CC diagnostic assays and for oxidative treatment of waste streams. NB. This
CC sequence is not given in the specification but was created from the wild
CC type sequence given by the inventors. The positions of the substitutions
CC given do not include the initiation codon
XX
SQ Sequence 154 AA;

Query Match 88.2%; Score 700; DB 2; Length 154;
Best Local Similarity 86.8%; Pred. No. 9.3e-70;
Matches 132; Conservative 9; Mismatches 11; Indels 0; Gaps 0;
QY 2 LSEGEQVLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDL 61
Db 3 LSDGEWQVNLVWVGKVEADIAHGQEVIRLFTGHPEITLTKDFEFKHLKTEAEKASEDL 62
QY 62 KKHGTVTLTALGAILKKKGHEALPLAQSHATKHKIPKYLEFTSEAIHVLHSHRHPG 121
Db 63 KKHGTVTLTALGAILKKKGHEALPLAQSHATKHKIPKYLEFTSEAIHVLHSHRHPG 122
QY 122 DFGADAQAGAMNKALELFRNDIAAKYKELGYQG 153
Db 123 DFGADAQAGAMTKALELFRNDIAAKYKELGYQG 154

RESULT 12
AAW29744
ID AAW29744 standard; protein; 154 AA.
XX
AC AAW29744;
XX
DT 26-OCT-1998 (first entry)
XX
DE Modified myoglobin protein 10.
XX
KW Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);
KW Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyse;
KW phenolic; polymer; antigen; antibody; diagnostic assay.
XX
OS Equus sp.
OS Synthetic.
XX
FN CA2185605-A.
XX
PD 17-MAR-1998.
XX
PF 16-SEP-1996; 96CA-02185605.
XX
PR 16-SEP-1996; 96CA-02185605.
XX
PA (MAUK/) MAUK A G.
PA (WANL/) WAN L.
PA (LEEH/) LEE H.
PA (BRAY/) BRAYER G D.
PA (TONG/) TONG H.
PA (SMIT/) SMITH M.
XX
PI Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;
XX
DR WPI; 1998-388655/34.
XX

PT New modified myoglobin with amino acid alterations has increased
PT peroxidase activity - used for catalysing oxidation of substrates with
PT peroxides, e.g. in waste treatment and as label for antigens.
XX
PS Claim 21; Page; 23pp; English.
XX
CC Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic
CC horse heart myoglobin (Mb) proteins, produced by random mutagenesis of
CC the synthetic Mb gene. This present sequence contains a substitution at
CC Lys45Glu, which has been found to enhance the peroxidase activity of the
CC protein in general, and it may also exhibit enhanced Mn(II) peroxidase
CC activity as well. This is due to the enhanced rate of reaction of Fe(III)

CC Mb with hydrogen peroxide and also because the substitution at residue 45
CC creates a new Mn binding site which is stronger than the natural Mn
CC binding site of Mb. This sequence also contains substitutions at
CC Thr39Ile, Phe46Leu, and Ile107Phe, which have been shown to enhance the
CC peroxidase activity of Mb. These protein can be used to catalyse oxidation
CC of a substrate by peroxidase, e.g. oxidation of Mn2+. Peroxidases are
CC used for oxidising phenolics to polymers, as labels for e.g. antigens,
CC antibodies or oligonucleotides, in diagnostic assays and for oxidative
CC treatment of waste streams. NB. This sequence is not given in the
CC specification but was created from the wild type sequence given by the
CC inventors. The positions of the substitutions given do not include the
CC initiation codon
XX
SQ Sequence 154 AA;

Query Match 88.0%; Score 699; DB 2; Length 154;
Best Local Similarity 86.2%; Pred. No. 1.2e-69;
Matches 131; Conservative 8; Mismatches 13; Indels 0; Gaps 0;
QY 2 LSEGEQVLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDL 61
Db 3 LSDGEWQVNLVWVGKVEADIAHGQEVIRLFTGHPEITLTKDFEFKHLKTEAEKASEDL 62
QY 62 KKHGTVTLTALGAILKKKGHEALPLAQSHATKHKIPKYLEFTSEAIHVLHSHRHPG 121
Db 63 KKHGTVTLTALGAILKKKGHEALPLAQSHATKHKIPKYLEFTSEAIHVLHSHRHPG 122
QY 122 DFGADAQAGAMNKALELFRNDIAAKYKELGYQG 153
Db 123 DFGADAQAGAMTKALELFRNDIAAKYKELGYQG 154

RESULT 13
ADC31106
ID ADC31106 standard; protein; 166 AA.
XX
AC ADC31106;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human novel polypeptide sequence, SEQ ID NO:1188.
XX
KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW gene therapy; chromosome 22q13.1.
XX
OS Homo sapiens.
XX
FN WO2003029271-A2.
XX
PD 10-APR-2003.
XX
PF 24-SEP-2002; 2002WO-US030474.
XX
PR 24-SEP-2001; 2001US-0324631P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Hailey-Vicente D, Drmanac RT;
XX
DR WPI; 2003-371981/35.
DR N-FSDB; ADC30135.
XX

PT New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or

Db 115 KKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIQVLQSKHPG 174
QY 122 DFGADAQGAMNKALELFRKDIAAKYKELGYQG 153
Db 175 DFGADAQGAMNKALELFRKDIAAKYKELGYQG 206
RESULT 15
AAW62273
ID AAW62273 standard; protein; 154 AA.
XX
AC AAW62273;
XX
DT 26-OCT-1998 (first entry)
XX
DE Modified myoglobin protein 4.
XX
KW Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);
KW Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyze;
KW phenolic; polymer; antigen; antibody; diagnostic assay.
XX
OS Equus sp.
OS Synthetic.
XX
PN CA2185605-A.
XX
PD 17-MAR-1998.
XX
PF 16-SEP-1996; 96CA-02185605.
XX
PR 16-SEP-1996; 96CA-02185605.
XX
PA (MAUK/) MAUK A G.
PA (WANL/) WAN L.
PA (LEEH/) LEE H.
PA (BRAY/) BRAYER G D.
PA (TONG/) TONG H.
PA (SMIT/) SMITH M.
XX
PI Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;
XX WPI; 1998-388655/34.
DR
XX
PT New modified myoglobin with amino acid alterations has increased
PT peroxidase activity - used for catalysing oxidation of substrates with
PT peroxides, e.g. in waste treatment and as label for antigens.
XX
PS Claim 23; Page; 23pp; English.
XX
CC Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic
CC horse heart myoglobin (Mb) proteins, produced by random mutagenesis of
CC the synthetic Mb gene. This present sequence contains a substitution at
CC Lys45Glu, which has been found to enhance the peroxidase activity of the
CC protein in general, and it may also exhibit enhanced Mn(II) peroxidase
CC activity as well. This is due to the enhanced rate of reaction of Fe(III)
CC Mb with hydrogen peroxide and also because the substitution at residue 45
CC creates a new Mn binding site which is stronger than the natural Mn
CC binding site of Mb. This protein also contains a substitution at
CC Lys83Glu, which was also found to be required for Mn binding, and
CC substitutions at His97Leu and Ser92Ala to increase the rate of Mn(II)
CC turnover. These protein can be used to catalyze oxidation of a substrate
CC by peroxidase, e.g. oxidation of Mn2+. Peroxidases are used for oxidising
CC phenolics to polymers, as labels for e.g. antigens, antibodies or
CC oligonucleotides, in diagnostic assays and for oxidative treatment of
CC waste streams. NB. This sequence is not given in the specification but
CC was created from the wild type sequence given by the inventors. The
CC positions of the substitutions given do not include the initiation codon
XX Sequence 154 AA;
SQ

Query Match 87.8%; Score 697; DB 2; Length 154;
Best Local Similarity 86.2%; Pred. No. 2e-69;
Matches 131; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLFKXSHPETLEKFDKFKHLKTEAEKASEDL 61
Db 3 LSDGEWQQVLNVWGKVEADIAGHGQEVLRIRLFTGHPETLEKFDKFKHLKTEAEKASEDL 62
QY 62 KKHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEATIHVLSRHPG 121
Db 63 KEHGTVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISDAIHVLSRHPG 122
QY 122 DFGADAQGAMNKALELFRKDIAAKYKELGYQG 153
Db 123 DFGADAQGAMTKALELFRNDIAAKYKELGYQG 154

Search completed: August 10, 2004, 15:26:14
Job time : 27.4879 secs

Blank

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	794	100.0	153	3	US-08-992-176-7	Sequence 7, Appli
2	778.5	98.0	152	3	US-08-992-176-10	Sequence 10, Appl
3	717	90.3	154	2	US-08-710-330A-11	Sequence 11, Appl
4	696	87.7	154	3	US-08-817-787-18	Sequence 18, Appl
5	584	73.6	113	2	US-08-710-330A-2	Sequence 2, Appli
6	525	66.1	113	2	US-08-710-330A-1	Sequence 1, Appli
7	515	64.9	113	2	US-08-710-330A-5	Sequence 5, Appli
8	505	63.6	113	2	US-08-710-330A-4	Sequence 4, Appli
9	504	63.5	113	2	US-08-710-330A-3	Sequence 3, Appli
10	240	30.2	60	3	US-08-817-787-22	Sequence 22, Appl
11	240	30.2	64	3	US-08-817-787-24	Sequence 24, Appl
12	213	26.8	57	3	US-08-817-787-20	Sequence 20, Appl
13	162	20.4	141	1	US-08-240-712-18	Sequence 18, Appl
14	162	20.4	141	1	US-08-443-890-18	Sequence 18, Appl
15	162	20.4	141	2	US-08-484-686B-65	Sequence 65, Appl
16	162	20.4	141	3	US-08-463-160B-66	Sequence 66, Appl
17	162	20.4	141	5	PCT-US92-09752-18	Sequence 18, Appl
18	152	19.1	141	3	US-09-058-562-18	Sequence 18, Appl
19	150	18.9	146	4	US-09-353-719-1	Sequence 1, Appli
20	140	17.6	141	2	US-08-627-173-21	Sequence 21, Appl
21	140	17.6	141	2	US-08-535-882A-21	Sequence 21, Appl
22	140	17.6	141	3	US-08-316-424A-7	Sequence 7, Appli
23	140	17.6	141	3	US-09-005-546-21	Sequence 21, Appl
24	140	17.6	141	4	US-08-477-669-7	Sequence 7, Appli
25	140	17.6	141	4	US-10-128-581-26	Sequence 26, Appl
26	135	17.0	146	2	US-08-619-708A-4	Sequence 4, Appli
27	128	16.1	141	2	US-08-627-173-19	Sequence 19, Appl

```
; CURRENT APPLICATION NUMBER: US/08/992,176
; CURRENT FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: JP 8-340727
; EARLIER FILING DATE: 1996-12-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 10
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: alpha-protein
; OTHER INFORMATION: (lmhc)
US-08-992-176-10

Query Match          98.0%; Score 778.5; DB 3; Length 152;
Best Local Similarity 99.3%; Pred. No. 1.1e-79;
Matches 152; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 VLSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKEDRPFKHLKTEAEKASEDL 60
Db 1 VLSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKEDRPFKHLKTEAEKASEDL 60
QY 61 LKKGVTVLTAALGAILKKKGHEAEKPLAQSHATKHKIPIKYLEFISEAIHVLHSRHP 120
Db 61 LKKGVTVLTAALGAILKKKGHEAEKPLAQSHATKHKIPIKYLEFISEAIHVLHS-HP 119
QY 121 GDFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
Db 120 GDFGADAQAGAMNKALELFRKDIAAKYKELGYQG 152

RESULT 3
US-08-710-330A-11
; Sequence 11, Application US/08710330A
; Patent No. 5854041
; GENERAL INFORMATION:
; APPLICANT: Brayer, Gary D.
; APPLICANT: Lee, Hung
; APPLICANT: Mauk, Grant A.
; APPLICANT: Smith, Michael
; APPLICANT: Tong, Harry
; APPLICANT: Wan, Lianglu
; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,330A
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GAY, David A.
; REGISTRATION NUMBER: 39,200
; REFERENCE/DOCKET NUMBER: P-SM 2262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 154 amino acids
; TYPE: amino acid
; STRANDEDNESS:
```

```
; TOPOLOGY: linear
US-08-710-330A-11

Query Match          90.3%; Score 717; DB 2; Length 154;
Best Local Similarity 88.2%; Pred. No. 9e-73;
Matches 134; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKEDRPFKHLKTEAEKASEDL 61
Db 3 LSDGEWQQLVNWGVKVEADIAHGQGVLRFTGHPETLEKFDKFKHLKTEAEKASEDL 62
QY 62 KKHGVTVLTAALGAILKKKGHEAEKPLAQSHATKHKIPIKYLEFISEAIHVLHSRHPG 121
Db 63 KKHGVTVLTAALGAILKKKGHEAEKPLAQSHATKHKIPIKYLEFISDAIHHVLSKHPG 122
QY 122 DFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
Db 123 DFGADAQAGAMTKALELFRNDIAAKYKELGFGQ 154

RESULT 4
US-08-817-787-18
; Sequence 18, Application US/08817787
; Patent No. 6294353
; GENERAL INFORMATION:
; APPLICANT: Pack, Peter
; APPLICANT: Lupas, Andrei
; TITLE OF INVENTION: TARGETED HETERO-ASSOCIATION OF
; TITLE OF INVENTION: RECOMBINANT PROTEINS TO MULTI-FUNCTIONAL COMPLEXES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,787
; FILING DATE: 23-SEP-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04117
; FILING DATE:
; PRIOR APPLICATION DATA: EP 94 11 6558.1
; FILING DATE: 20-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MORPHO/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 154 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-08-817-787-18

Query Match          87.7%; Score 696; DB 3; Length 154;
Best Local Similarity 84.2%; Pred. No. 2e-70;
Matches 128; Conservative 14; Mismatches 10; Indels 0; Gaps 0;
```

QY 2 LSEGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKHLEKTEAEKASD 61
Db 3 LSGEQLVNLVWVGKVEADIPRGQGVLRIRLFKSHPETLEKDFRKHLEKSEDEMKASD 62
QY 62 KKHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEAIIHVLHSHRHPG 121
Db 63 KKHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEAIIHVLHSHRHPG 122
QY 122 DFCADAGGAMKALELFRKDIAKYKELGVQ 153
Db 123 DFCADAGGAMKALELFRKDIAKYKELGVQ 154

RESULT 5

US-08-710-330A-2
; Sequence 2, Application US/08710330A
; Patent No. 5854041
; GENERAL INFORMATION:
; APPLICANT: Brayer, Gary D.
; APPLICANT: Lee, Hung
; APPLICANT: Mauk, Grant A.
; APPLICANT: Smith, Michael
; APPLICANT: Tong, Harry
; APPLICANT: Wan, Lianglu
; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,330A
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GAY, David A.
; REGISTRATION NUMBER: 39,200
; REFERENCE/DOCKET NUMBER: P-SM 2262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-710-330A-2

Query Match 73.6%; Score 584; DB 2; Length 113;
Best Local Similarity 100.0%; Pred. No. 5e-58;
Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKHLEKTEAEKASD 60
Db 1 VLSEGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKHLEKSEDEMKASD 60
QY 61 LKKHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEAIIH 113
Db 61 LKKHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEAIIH 113

RESULT 6

US-08-710-330A-1

; Sequence 1, Application US/08710330A
; Patent No. 5854041
; GENERAL INFORMATION:
; APPLICANT: Brayer, Gary D.
; APPLICANT: Lee, Hung
; APPLICANT: Mauk, Grant A.
; APPLICANT: Smith, Michael
; APPLICANT: Tong, Harry
; APPLICANT: Wan, Lianglu
; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,330A
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GAY, David A.
; REGISTRATION NUMBER: 39,200
; REFERENCE/DOCKET NUMBER: P-SM 2262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-710-330A-1

Query Match 66.1%; Score 525; DB 2; Length 113;
Best Local Similarity 87.5%; Pred. No. 2.1e-51;
Matches 98; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKHLEKTEAEKASD 61
Db 2 LSGEQLVNLVWVGKVEADIPRGQGVLRIRLFKSHPETLEKDFRKHLEKSEDEMKASD 61
QY 62 KKHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEAIIH 113
Db 62 KKHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEAIIH 113

RESULT 7

US-08-710-330A-5
; Sequence 5, Application US/08710330A
; Patent No. 5854041
; GENERAL INFORMATION:
; APPLICANT: Brayer, Gary D.
; APPLICANT: Lee, Hung
; APPLICANT: Mauk, Grant A.
; APPLICANT: Smith, Michael
; APPLICANT: Tong, Harry
; APPLICANT: Wan, Lianglu
; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California


```

Db      2 LSDGEQAVLNAMQKVADVAGHQEVLRITFGHPETLEKFKFKHLKTEAEMKASDL 61
QY      62 KHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIH 113
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      62 KHGNTVLTALGGILKKKGHHEAELKPLAQSHATKHKIPKYLEFISDAIIH 113
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
US-08-817-787-22
; Sequence 22, Application US/08817787
; Patent No. 6294353
; GENERAL INFORMATION:
; APPLICANT: Pack, Peter
; APPLICANT: Lupas, Andrei
; TITLE OF INVENTION: TARGETED HETERO-ASSOCIATION OF
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,787
; FILING DATE: 23-SEP-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04117
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94 11 6558.1
; FILING DATE: 20-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MORPHO/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 424:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-817-787-22

Query Match      30.2%; Score 240; DB 3; Length 60;
Best Local Similarity 90.2%; Pred. No. 8.6e-20;
Matches 46; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      48 HLKTEAEMKASDLKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHK 98
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      3 HLKSEDEMKASDLKKHGATVLTALGGILKKKGHHEAELKPLAQSHATKHK 53
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
US-08-817-787-24
; Sequence 24, Application US/08817787
; Patent No. 6294353
; GENERAL INFORMATION:
; APPLICANT: Pack, Peter
; APPLICANT: Lupas, Andrei
; TITLE OF INVENTION: TARGETED HETERO-ASSOCIATION OF
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,787
; FILING DATE: 23-SEP-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04117
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94 11 6558.1
; FILING DATE: 20-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MORPHO/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-817-787-22

Query Match      30.2%; Score 240; DB 3; Length 60;
Best Local Similarity 90.2%; Pred. No. 8.6e-20;
Matches 46; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      48 HLKTEAEMKASDLKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHK 98
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      3 HLKSEDEMKASDLKKHGATVLTALGGILKKKGHHEAELKPLAQSHATKHK 53
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-08-817-787-20
; Sequence 20, Application US/08817787
; Patent No. 6294353
; GENERAL INFORMATION:
; APPLICANT: Pack, Peter
; APPLICANT: Lupas, Andrei
; TITLE OF INVENTION: TARGETED HETERO-ASSOCIATION OF
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,787
; FILING DATE: 23-SEP-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04117
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94 11 6558.1
; FILING DATE: 20-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MORPHO/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 64 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-817-787-24

Query Match      30.2%; Score 240; DB 3; Length 64;
Best Local Similarity 80.0%; Pred. No. 9.4e-20;
Matches 44; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY      99 IPIKYLEFISEAIIHVLHSHRHPGDFGADAGGAMNKALELFRKDIKAKYKELGVQG 153
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      3 IPIKVEFISECIIQVLSKHPGDFGADAGGAMNKALELFRKDMASNYKELGVQG 57
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-08-817-787-20
; Sequence 20, Application US/08817787
; Patent No. 6294353
; GENERAL INFORMATION:
; APPLICANT: Pack, Peter
; APPLICANT: Lupas, Andrei
; TITLE OF INVENTION: TARGETED HETERO-ASSOCIATION OF
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,787
; FILING DATE: 23-SEP-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04117
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94 11 6558.1
; FILING DATE: 20-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MORPHO/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 64 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-817-787-24

```

```

; MOLECULE TYPE: peptide
US-08-240-712-18

Query Match 20.4%; Score 162; DB 1; Length 141;
Best Local Similarity 28.1%; Pred. NO. 1.6e-10;
Matches 41; Conservative 29; Mismatches 70; Indels 6; Gaps 1;

QY 2 LSEGEWLVHVNKVEADVAGHGQDILIRLFKSHPETLEKDFRKHKTAEAMKASEDL 61
Db 2 LTKTERTIIVSMWAKISTOADTIGTETLERLFLSHPQTKTYFFPHFDLHGPSAQIRA---- 57

QY 62 KHGVTVLTALGAILKKKGHEAEALKPLAQSHATKHKIPIKYLEFISEAIIHVLHGRHPG 121
Db 58 --HGKVVAAVGDAVKSIDDIGALSLSLSELHAYILRVPVNFKLSHCLLVTLAARFPA 115

QY 122 DFGADAQAGMNALELFRKDIAKYK 147
Db 116 DFTAEAAAWDKFLSVSVSLTEKYR 141

RESULT 14
US-08-443-890-18
; Sequence 18, Application US/08443890
; Patent No. 5739011
; GENERAL INFORMATION:
; APPLICANT: ANDERSON, DAVID C.
; APPLICANT: MATHEWS, ANTONY JAMES
; APPLICANT: STETLER, GARY L.
; TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC
; TITLE OF INVENTION: HEMOGLOBINS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,890
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/240,712
; FILING DATE: 09-MAY-1994
; APPLICATION NUMBER: PCT/US92/09752
; FILING DATE: 13-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, IVER P
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: ANDERSON-6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-443-890-18

Query Match 20.4%; Score 162; DB 1; Length 141;
Best Local Similarity 28.1%; Pred. NO. 1.6e-10;
Matches 41; Conservative 29; Mismatches 70; Indels 6; Gaps 1;

```


QY	2	LSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFORFKHLKTEAEMKASEDL	61
Db	2	LTKTERTIIVSWAKISTAQDTIGTETLERLFLSHPQTKEYFPHFDLHPSAQIR-----	57
QY	62	KKHGVTVLTALGAILKKKGHEAELKPIAQSHATKHKIPYKLEFISEAIHVLHSHRHP	121
Db	58	--HGSKVAAGVDGVKSIDDIGGALSKLSHVAIIRVDPVNFKLLSHCLLVTLAARFFA	115
QY	122	DFGADAQGMANKALELFRKDIAAYK	147
Db	116	DFAEAHAADWKFLSVGSSVITEYR	141

RESULT 15

US-08-484-686B-65
; Sequence 65, Application US/08484686B
; Patent No. 5827693
; GENERAL INFORMATION:
; APPLICANT: De Angelo, Joseph
; APPLICANT: Motwani, Nalini
; APPLICANT: Bajwa, Wajeeh
; TITLE OF INVENTION: Expression of Recombinant Hemoglobin and
; TITLE OF INVENTION: Hemoglobin Variants in Yeast
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,686B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/368,407
; FILING DATE: 29-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/876,290
; FILING DATE: 29-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/684,611
; FILING DATE: 12/APR/1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Flintoft, Gerald J.
; REGISTRATION NUMBER: 20,823
; REFERENCE/DOCKET NUMBER: 6666-043-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-484-686B-65

```
Query Match      20.4%; Score 162; DB 2; Length 141;
Best Local Similarity 28.1%; Pred. No. 1.6e-10;
Matches 4; Conservative 29; Mismatches 70; Indels 6; Gaps 1;

2 LSEGEVLVHVWAKVEADVAGHQDILIRLPKSPHPTLEKFDRPKHLKTEAEKASEDL 61
| : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
LTKTERTIIVSWMAKISTQADTIGTETLERFLPSHPQTGYFPHFDLHGSAQLRA --- 57
```

Qy	62	KKGVTVTALGAILKKGKHHAEKPLAQSHATKHKIPIKYLBFI	121
Db	58	--HGSKVVAAVGAIVSIDDIGGALSLSLSELHAYILRVDPVNFLLSHCLLVTAAARFPA	115
Qy	122	DFGADAGAMNKALELFRKIDIAKYK	147
Db	116	DFTAEEAHAADKFLSVVSSVLTKEYR	141

Search completed: August 10, 2004, 15:30:20
Job time : 8.77966 secs

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:28:46 ; Search time 22.2276 Seconds
(without alignments)
2159.184 Million cell updates/sec

Title: US-09-455-978b-76
Perfect score: 794
Sequence: 1 VLSEGEQLVHLVWAKVEAD.....ALELFRKDIAAKYKELGYQG 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 segs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	696	87.7	154	16	US-10-408-765A-106	Sequence 106, Appl
2	162	20.4	141	9	US-09-977-577-19	Sequence 19, Appl
3	152	19.1	31	15	US-10-289-009-17	Sequence 17, Appl
4	140	17.6	141	9	US-09-839-164-7	Sequence 7, Appl
5	140	17.6	141	14	US-10-128-581-26	Sequence 26, Appl
6	140	17.6	141	16	US-10-463-699-28	Sequence 28, Appl
7	133	16.8	31	15	US-10-289-009-4	Sequence 4, Appl
8	133	16.8	154	12	US-10-378-029-102	Sequence 102, Appl
9	132	16.6	142	12	US-10-424-599-227247	Sequence 227247,
10	132	16.6	142	14	US-10-205-219-189	Sequence 189, Appl
11	128	16.1	141	9	US-09-839-164-5	Sequence 5, Appl
12	128	16.1	141	14	US-10-128-581-24	Sequence 24, Appl
13	128	16.1	141	16	US-10-463-699-26	Sequence 26, Appl
14	127	16.0	146	9	US-09-839-164-8	Sequence 8, Appl
15	127	16.0	146	14	US-10-128-581-27	Sequence 27, Appl

16	127	15.0	146	16	US-10-463-699-29	Sequence 29, Appl
17	125	15.7	146	9	US-09-977-577-20	Sequence 20, Appl
18	124	15.6	146	9	US-09-977-577-17	Sequence 17, Appl
19	120	15.1	146	9	US-09-839-164-6	Sequence 6, Appl
20	120	15.1	146	14	US-10-128-581-25	Sequence 25, Appl
21	120	15.1	146	16	US-10-463-699-27	Sequence 27, Appl
22	114	14.4	147	9	US-09-147-490-8	Sequence 8, Appl
23	112	14.1	141	9	US-09-839-164-2	Sequence 2, Appl
24	112	14.1	141	9	US-09-977-577-14	Sequence 14, Appl
25	112	14.1	141	9	US-09-977-577-18	Sequence 18, Appl
26	112	14.1	141	12	US-10-280-725B-8	Sequence 8, Appl
27	112	14.1	141	13	US-10-085-853-31	Sequence 31, Appl
28	112	14.1	141	14	US-10-128-581-21	Sequence 21, Appl
29	112	14.1	141	14	US-10-280-679B-8	Sequence 8, Appl
30	112	14.1	141	16	US-10-463-699-23	Sequence 23, Appl
31	112	14.1	142	15	US-10-435-666-7	Sequence 7, Appl
32	111	14.0	141	16	US-10-408-765A-310	Sequence 310, Appl
33	104	13.1	122	15	US-10-115-482-52	Sequence 52, Appl
34	103	13.0	146	9	US-09-839-164-4	Sequence 4, Appl
35	103	13.0	146	9	US-09-977-577-15	Sequence 15, Appl
36	103	13.0	146	13	US-10-085-853-33	Sequence 33, Appl
37	103	13.0	146	14	US-10-128-581-23	Sequence 23, Appl
38	103	13.0	146	16	US-10-463-699-25	Sequence 25, Appl
39	103	13.0	146	16	US-10-408-765A-3019	Sequence 3019, Appl
40	101	12.7	19	15	US-10-289-009-12	Sequence 12, Appl
41	101	12.7	20	12	US-10-103-395-189	Sequence 189, Appl
42	100	12.6	147	9	US-09-977-577-21	Sequence 21, Appl
43	99	12.5	146	9	US-09-977-577-16	Sequence 16, Appl
44	98	12.3	146	16	US-10-408-765A-84	Sequence 84, Appl
45	96	12.1	18	15	US-10-289-009-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1

US-10-408-765A-106
; Sequence 106, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-106

Query Match	87.7%	Score 696;	DB 16;	Length 154;
Best Local Similarity	84.2%	Pred. No. 2.3e-69;		
Matches 128;	Conservative 14;	Mismatches 10;	Indels 0;	Gaps 0;
QY	2	LSGEQQLVHLVWAKVEADVAGHGDILIRLFKSHPETLEKFDPRFKHLKTEAKMKASDGL 61		
Db	3	LSDCGEQLVNLVWGKVEADIPGHGOEVLIRLFKSHPETLEKFDPRFKHLKSEDEMKASDGL 62		
QY	62	KKHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHGKTIPIKYLEFISEAIIHVLHSRHPG 121		
Db	63	KKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHGKTIPIKYLEFISECIIQVLQSKHPG 122		
QY	122	DFGADAQAGMNALELFRKDIAAKYKELGYQG 153		


```
RESULT 5
US-10-581-26
; Sequence 26, Application US/10128581
; Publication No. US20030104984A1
; GENERAL INFORMATION:
; APPLICANT: TSYRLOVA, IRENA
; WOLPE, STEPHEN D.
; TITLE OF INVENTION: HEMOGLOBIN ALPHA CHAIN PEPTIDE FRAGMENTS USEFUL
; FOR INHIBITING STEM CELL PROLIFERATION
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 MB disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/128,581
; FILING DATE: 24-APR-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/617,840
; FILING DATE: 17-JUL-2000
; APPLICATION NUMBER: US 09/005,546
; FILING DATE: 12-JAN-1998
; APPLICATION NUMBER: US 08/535,882
; FILING DATE: 28-SEP-1995
; APPLICATION NUMBER: US 08/316,424
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: PCT/US94/03349
; FILING DATE: 29-MAR-1994
; APPLICATION NUMBER: US 08/040,942
; FILING DATE: 31-MAR-1993
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-10-128-581-26
Query Match 17.6%; Score 140; DB 14; Length 141;
Best Local Similarity 27.6%; Pred. No. 2.1e-07;
Matches 42; Conservative 20; Mismatches 74; Indels 16; Gaps 4;
QY 1 VLSGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKPDFR--KHLKTEAEKAS 58
Db 1 VLSAADKANVKAAGKVGQAGAHGAELRMFLGPTTKTYFPHNLSH-----GS 52
QY 59 EDLKKHGVTVTALGAILKKKGHHE---AELKPLAQSHATKHKIPKYLEFISEAIIHVL 115
Db 53 DOVKAGQKVADAL---TKAVGHLDLPGALSALSDLHAHLKLRVDPVNFKLKSHCLLVTL 109
QY 116 HSRHPDGFADAGQAGAMNKALELFRKDIAAKYK 147
Db 110 AAHPDDPNPSVHASLDKFLANVSTVLTSKYR 141
RESULT 6
US-10-463-699-28
; Sequence 28, Application US/10463699
; Publication No. US20040081640A1
; GENERAL INFORMATION:
; APPLICANT: TSYRLOVA, IRENA
; WOLPE, STEPHEN D.
; TITLE OF INVENTION: HEMOGLOBIN ALPHA CHAIN PEPTIDE FRAGMENTS USEFUL
; FOR INHIBITING STEM CELL PROLIFERATION
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/128,581
; FILING DATE: 24-APR-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/617,840
; FILING DATE: 17-JUL-2000
; APPLICATION NUMBER: US 09/005,546
; FILING DATE: 12-JAN-1998
; APPLICATION NUMBER: US 08/535,882
; FILING DATE: 28-SEP-1995
; APPLICATION NUMBER: US 08/316,424
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: PCT/US94/03349
; FILING DATE: 29-MAR-1994
; APPLICATION NUMBER: US 08/040,942
; FILING DATE: 31-MAR-1993
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-10-128-581-26
Query Match 17.6%; Score 140; DB 14; Length 141;
Best Local Similarity 27.6%; Pred. No. 2.1e-07;
Matches 42; Conservative 20; Mismatches 74; Indels 16; Gaps 4;
QY 1 VLSGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKPDFR--KHLKTEAEKAS 58
Db 1 VLSAADKANVKAAGKVGQAGAHGAELRMFLGPTTKTYFPHNLSH-----GS 52
QY 59 EDLKKHGVTVTALGAILKKKGHHE---AELKPLAQSHATKHKIPKYLEFISEAIIHVL 115
Db 53 DOVKAGQKVADAL---TKAVGHLDLPGALSALSDLHAHLKLRVDPVNFKLKSHCLLVTL 109
QY 116 HSRHPDGFADAGQAGAMNKALELFRKDIAAKYK 147
Db 110 AAHPDDPNPSVHASLDKFLANVSTVLTSKYR 141
RESULT 7
US-10-289-009-4
; Sequence 4, Application US/10289009
; Publication No. US20030228700A1
; GENERAL INFORMATION:
; APPLICANT: Peters, Eric C.
; APPLICANT: Brock, Ansgar
; APPLICANT: Ericson, Christer
; APPLICANT: IRM LLC
; TITLE OF INVENTION: Labeling Reagent and Methods of Use
; FILE REFERENCE: 021288-00023005
; WOLPE, STEPHEN D.
; TITLE OF INVENTION: HEMOGLOBIN ALPHA CHAIN PEPTIDE FRAGMENTS
; USEFUL FOR INHIBITING STEM CELL PROLIFERATION
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 MB disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/463,699
; FILING DATE: 18-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 10/128,581
; FILING DATE: 24-APR-2002
; APPLICATION NUMBER: US 09/617,840
; FILING DATE: 17-JUL-2000
; APPLICATION NUMBER: US 09/005,546
; FILING DATE: 12-JAN-1998
; APPLICATION NUMBER: US 08/535,882
; FILING DATE: 28-SEP-1995
; APPLICATION NUMBER: US 08/316,424
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: PCT/US94/03349
; FILING DATE: 29-MAR-1994
; APPLICATION NUMBER: US 08/040,942
; FILING DATE: 31-MAR-1993
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-10-463-699-28
Query Match 17.6%; Score 140; DB 16; Length 141;
Best Local Similarity 27.6%; Pred. No. 2.1e-07;
Matches 42; Conservative 20; Mismatches 74; Indels 16; Gaps 4;
QY 1 VLSGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKPDFR--KHLKTEAEKAS 58
Db 1 VLSAADKANVKAAGKVGQAGAHGAELRMFLGPTTKTYFPHNLSH-----GS 52
QY 59 EDLKKHGVTVTALGAILKKKGHHE---AELKPLAQSHATKHKIPKYLEFISEAIIHVL 115
Db 53 DOVKAGQKVADAL---TKAVGHLDLPGALSALSDLHAHLKLRVDPVNFKLKSHCLLVTL 109
QY 116 HSRHPDGFADAGQAGAMNKALELFRKDIAAKYK 147
Db 110 AAHPDDPNPSVHASLDKFLANVSTVLTSKYR 141
```

```
; CURRENT APPLICATION NUMBER: US/10/289,009
; CURRENT FILING DATE: 2003-04-01
; PRIOR APPLICATION NUMBER: US 60/332,988
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 60/385,835
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 60/410,382
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: equine
; OTHER INFORMATION: myoglobin tryptic polypeptide #3
US-10-289-009-4

Query Match          16.8%; Score 133; DB 15; Length 31;
Best Local Similarity 76.7%; Pred. No. 1.6e-07;
Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 LSGEWQLVHVWAKVEADVAGHGQDILIR 31
Db 2 LSDGEWQVLNVGWKVEADVAGHGQEVLR 31

RESULT 8
US-10-378-029-102
; Sequence 102, Application US/10378029
; Publication No. US20040014087A1
; GENERAL INFORMATION:
; APPLICANT: HODGSON, David M.; LINCOLN, Stephen E.
; APPLICANT: RUSSO, Frank D.; SPIRO, Peter A.
; APPLICANT: BANVILLE, Steve C.; BRATCHER, Shawn R.
; APPLICANT: DUFOUR, Gerard E.; COHEN, Howard J.
; APPLICANT: ROSEN, Bruce; CHALUP, Michael S.
; APPLICANT: JACKSON, Jennifer L.; JONES, Annisa L.
; APPLICANT: YU, Jimmy Y.; GREENAWALT, Lila B.
; APPLICANT: PANZER, Scott R.; ROSEBERRY LINCOLN, Ann M.
; APPLICANT: WRIGHT, Rachel J.; DANIELS, Susan E.
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PT-1022-1 CIP
; CURRENT APPLICATION NUMBER: US/10/378,029
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 09/980,285
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: PCT/US00/15404
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: US 60/147,500
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,542
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,541
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,824
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,547
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,530
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,536
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,520
; PRIOR FILING DATE: 1999-08-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 104
; SOFTWARE: PERL Program
; SEQ ID NO 102
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040014087A1 441779.1.j.orf1
US-10-378-029-102

Query Match          16.8%; Score 133; DB 15; Length 154;
Best Local Similarity 26.7%; Pred. No. 1.4e-06;
Matches 40; Conservative 24; Mismatches 74; Indels 12; Gaps 3;

QY 1 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLPSHPETLEKFDREKHLKTEAEWKASED 60
Db 14 VLSAADKTNIKNCWGKIGGEGEALQRMFAFPPTTKTYFSDHIDVSPGSAQVKA--- 70
QY 61 LKKHGVTVLTALGAILKKKGHE---AELKPLAQSHATKHKIPIKYLEFISEAIIHVLS 117
Db 71 ---HGKKVADALA---KAADHVEDLPGALSTLSDLHAHKLKRVDPVNFKLSHCLLVTLAC 124
QY 118 RHPGDFGADAQGMNKALELFRKDIKAAKYK 147
Db 125 HHPGDFTPAMHASLDKFLASVSTVLTSKYR 154

RESULT 9
US-10-424-599-227247
; Sequence 227247, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 227247
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_47233C.1.pep
US-10-424-599-227247

Query Match          16.6%; Score 132; DB 12; Length 142;
Best Local Similarity 26.7%; Pred. No. 1.7e-06;
Matches 40; Conservative 24; Mismatches 74; Indels 12; Gaps 3;

QY 1 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLPSHPETLEKFDREKHLKTEAEWKASED 60
Db 2 VLSADDKTNIKNCWGKIGGEGEALQRMFAFPPTTKTYFSDHIDVSPGSAQVKA--- 58
QY 61 LKKHGVTVLTALGAILKKKGHE---AELKPLAQSHATKHKIPIKYLEFISEAIIHVLS 117
Db 59 ---HGKKVADALA---KAADHVEDLPGALSTLSDLHAHKLKRVDPVNFKLSHCLLVTLAC 112
QY 118 RHPGDFGADAQGMNKALELFRKDIKAAKYK 147
Db 113 HHPGDFTPAMHASLDKFLASVSTVLTSKYR 142

RESULT 10
US-10-205-219-189
; Sequence 189, Application US/10205219
; Publication No. US20030138803A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
; APPLICANT: Brooksbank, Robert
; APPLICANT: Pincock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
```

```
; FILE REFERENCE: WL-A-018200
; CURRENT APPLICATION NUMBER: US/10/205,219
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 189
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Globin, alpha, major
US-10-205-219-189

Query Match          16.6%; Score 132; DB 14; Length 142;
Best Local Similarity 26.7%; Pred. No. 1.7e-06;
Matches 40; Conservative 24; Mismatches 74; Indels 12; Gaps 3;

QY 1 VLSEGEWQLVHVWAKVADVAGHGODILIRLFKSHPETLEKFDKFRF--KHLKTEAEAMKASD 60
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 VLSADDTNKNCKWGKIGGCGEGEALQRMFAFPPTTKTFESHIDVSPGSAQVKA--- 58

QY 61 LKHGVTULTALGAILKKKGHE---AELKPLAQSHATKHKIPKYLEFISEAIIHVLHS 117
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 ---HGKQVADALA---KAADHVEDLPGLSTLSDLHAHKLKRVDPVNFKFLSHCLLVTLAC 112

QY 118 RHPGDFGADAQAGAMNKALELFRKIDIAAKYK 147
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 113 HHPGDFTPAMHASLDKFLASVSVTLTSKYR 142

RESULT 11
US-09-839-164-5
; Sequence 5, Application US/09839164
; Patent No. US20020098583A1
; GENERAL INFORMATION:
; APPLICANT: KOZLOV, VLADIMIR
; TSYRLOVA, IRENA
; TITLE OF INVENTION: INHIBITOR OF STEM CELL PROLIFERATION AND
; USES THEREOF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8th FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/839,164
; FILING DATE: 23-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,668
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/316,424
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: PCT/US94/03349
; FILING DATE: 29-MAR-1994
; APPLICATION NUMBER: US 08/040,942
; FILING DATE: 31-MAR-1993
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-205-219-189

Query Match          16.6%; Score 132; DB 14; Length 141;
Best Local Similarity 27.0%; Pred. No. 4.6e-06;
Matches 41; Conservative 22; Mismatches 73; Indels 16; Gaps 4;

QY 1 VLSEGEWQLVHVWAKVADVAGHGODILIRLFKSHPETLEKFDKFRF--KHLKTEAEAMKAS 58
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 VLSGEDKSNIAKAWGKIGGCGAEGAEALERMFASFPPTTKTFPHFDVSH-----GS 52

QY 59 EDLKKHGVTLTALGAILKKKGHE---AELKPLAQSHATKHKIPKYLEFISEAIIHVL 115
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 AQVKGHGKQVADALAS---AAGHLDDLPGLSALSLSDLHAHKLKRVDPVNFKLLSHCLLVTL 109

QY 116 HSRHFGDGAQAGAMNKALELFRKIDIAAKYK 147
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 110 ASHHPADFTPAVHASLDKFLASVSVTLTSKYR 141

RESULT 12
US-10-128-581-24
; Sequence 24, Application US/10128581
; Publication No. US20030104984A1
; GENERAL INFORMATION:
; APPLICANT: TSYRLOVA, IRENA
; WOLPE, STEPHEN D.
; TITLE OF INVENTION: HEMOGLOBIN ALPHA CHAIN PEPTIDE FRAGMENTS USEFUL
; FOR INHIBITING STEM CELL PROLIFERATION
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/128,581
; FILING DATE: 24-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/617,840
; FILING DATE: 17-JUL-2000
; APPLICATION NUMBER: US 09/005,546
; FILING DATE: 12-JAN-1998
; APPLICATION NUMBER: US 08/535,882
; FILING DATE: 28-SEP-1995
; APPLICATION NUMBER: US 08/316,424
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: PCT/US94/03349
; FILING DATE: 29-MAR-1994
; APPLICATION NUMBER: US 08/040,942
; FILING DATE: 31-MAR-1993
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-10-128-581-24

Query Match          16.1%; Score 128; DB 14; Length 141;
Best Local Similarity 27.0%; Pred. No. 4.6e-06;
Matches 41; Conservative 22; Mismatches 73; Indels 16; Gaps 4;
```


Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:26:21 ; Search time 94.2821 Seconds
(without alignments)
1583.930 Million cell updates/sec

Title: US-09-455-978B-76
Perfect score: 794
Sequence: 1 VLSEGEWQLVHLVWAKVEAD.....ALELFRKDIAKYKELGYQG 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main:*

- 1: /cgn2_6/ptodata/2/paa/PTCUS_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
- 22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
- 23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
- 24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
- 25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
- 26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
- 27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
- 28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
- 29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
- 30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
- 31: /cgn2_6/ptodata/2/paa/US106_COMB.pep.*
- 32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
- 33: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID Description

1	794	100.0	153	18	US-09-455-978B-76	Sequence 76, Appli
2	794	100.0	153	21	US-09-700-708-4	Sequence 4, Appli
3	794	100.0	154	22	US-09-791-537-51499	Sequence 51499, A
4	789	99.4	153	22	US-09-791-537-41444	Sequence 41444, A
5	789	99.4	154	22	US-09-791-537-23678	Sequence 23678, A
6	788	99.2	154	22	US-09-791-537-13857	Sequence 13857, A
7	788	99.2	154	22	US-09-791-537-22569	Sequence 22569, A
8	786	99.0	154	22	US-09-791-537-23088	Sequence 23088, A
9	786	99.0	154	22	US-09-791-537-23776	Sequence 23776, A
10	786	99.0	154	22	US-09-791-537-102872	Sequence 102872, A
11	785	98.9	154	22	US-09-791-537-22535	Sequence 22535, A
12	785	98.9	154	22	US-09-791-537-23775	Sequence 23775, A
13	785	98.9	154	22	US-09-791-537-102151	Sequence 102151, A
14	785	98.9	154	22	US-09-791-537-102898	Sequence 102898, A
15	784	98.7	153	22	US-09-791-537-51498	Sequence 51498, A
16	784	98.7	154	22	US-09-791-537-13855	Sequence 13855, A
17	784	98.7	154	22	US-09-791-537-17232	Sequence 17232, A
18	784	98.7	154	22	US-09-791-537-22555	Sequence 22555, A
19	784	98.7	154	22	US-09-791-537-102870	Sequence 102870, A
20	783	98.6	151	22	US-09-791-537-71491	Sequence 71491, A
21	783	98.6	154	22	US-09-791-537-20784	Sequence 20784, A
22	783	98.6	154	22	US-09-791-537-33391	Sequence 33391, A
23	783	98.6	154	22	US-09-791-537-74339	Sequence 74339, A
24	783	98.6	154	22	US-09-791-537-102153	Sequence 102153, A
25	783	98.6	154	22	US-09-791-537-102916	Sequence 102916, A
26	783	98.6	154	22	US-09-791-537-151603	Sequence 151603, A
27	782	98.5	154	22	US-09-791-537-20811	Sequence 20811, A
28	782	98.5	154	22	US-09-791-537-73835	Sequence 73835, A
29	782	98.5	154	22	US-09-791-537-102896	Sequence 102896, A
30	781	98.4	154	22	US-09-791-537-13847	Sequence 13847, A
31	781	98.4	154	22	US-09-791-537-23671	Sequence 23671, A
32	781	98.4	154	22	US-09-791-537-102150	Sequence 102150, A
33	780	98.2	154	22	US-09-791-537-67801	Sequence 67801, A
34	780	98.2	154	22	US-09-791-537-102163	Sequence 102163, A
35	779	98.1	154	22	US-09-791-537-23158	Sequence 23158, A
36	779	98.1	154	22	US-09-791-537-23675	Sequence 23675, A
37	779	98.1	154	22	US-09-791-537-27099	Sequence 27099, A
38	778	98.0	151	22	US-09-791-537-108163	Sequence 108163, A
39	778	98.0	154	22	US-09-791-537-23159	Sequence 23159, A
40	778	98.0	154	22	US-09-791-537-23676	Sequence 23676, A
41	778	98.0	154	22	US-09-791-537-102159	Sequence 102159, A
42	775	97.6	154	22	US-09-791-537-23089	Sequence 23089, A
43	773	97.4	153	22	US-09-791-537-9291	Sequence 9291, Ap
44	772	97.2	154	22	US-09-791-537-27073	Sequence 27073, A
45	772	97.2	154	22	US-09-791-537-77660	Sequence 77660, A

ALIGNMENTS

RESULT 1
US-09-455-978B-76
; Sequence 76, Application US/09455978B
; GENERAL INFORMATION:
; APPLICANT: Alam, Magsudul
; APPLICANT: Larsen, Randy
; TITLE OF INVENTION: HEME PROTEINS HEMAT-HS AND HEMAT-BS AND THEIR USE IN
; FILE REFERENCE: 201040/1020
; CURRENT APPLICATION NUMBER: US/09/455,978B
; CURRENT FILING DATE: 1999-12-06
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 76
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Sperm-whale myoglobin
US-09-455-978B-76

Query Match 100.0%; Score 794; DB 18; Length 153;
Best Local Similarity 100.0%; Pred. No. 2.4e-77;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

RESULT 2

```

US-09-700-708-4
; Sequence 4, Application US/09700708
; GENERAL INFORMATION:
; APPLICANT: ITAI, Akiko
; APPLICANT: ITAI, Reiko
; APPLICANT: TOMIOKA, Nobuo
; TITLE OF INVENTION: Method For Predicting Functions of Protein
; FILE REFERENCE: P20234
; CURRENT APPLICATION NUMBER: US/09/700,708
; CURRENT FILING DATE: 2000-11-24
; PRIOR APPLICATION NUMBER: PCT/JP98/02302
; PRIOR FILING DATE: 1998-05-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 153
; TYPE: prt
; ORGANISM: Whale
US-09-700-708-4

```

RESULT 3

```

US-09-791-537-51499
; Sequence 51499, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAM
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 51499
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb 1DUKA
US-09-791-537-51499

```

RESULT 5

```

US-09-791-537-23678
; Sequence 23678, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAL
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 133055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23678
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb 2MGK
US-09-791-537-23678

```

RESULT 4

```

US-09-791-537-41444
; Sequence 41444, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biocomix, Inc.
; APPLICANT: Danzer, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURE
; FILE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 41444
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Physter catodon
US-09-791-537-41444

```

RESULT 5

```

US-09-791-537-23678
; Sequence 23678, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAL
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 133055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23678
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb 2MGK
US-09-791-537-23678

```

Best Local Similarity 99.3%; Pred. No. 8.5e-77;
Matches 152; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 60
Db 2 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 61
QY 61 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
Db 62 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 121
QY 121 GDFGADAQAGMANKALELFRKDIAAAYKELGYQG 153
Db 122 GNFGADAQAGMANKALELFRKDIAAAYKELGYQG 154

RESULT 6

US-09-791-537-13857
; Sequence 13857, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13857
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb ICIOA
US-09-791-537-13857

Query Match 99.2%; Score 788; DB 22; Length 154;
Best Local Similarity 98.7%; Pred. No. 1.1e-76;
Matches 151; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 60
Db 2 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 61
QY 61 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
Db 62 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 121
QY 121 GDFGADAQAGMANKALELFRKDIAAAYKELGYQG 153
Db 122 GNFGADAQAGMANKALELFRKDIAAAYKELGYQG 154

RESULT 7

US-09-791-537-22569
; Sequence 22569, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22569
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb IMLM
US-09-791-537-22569

Query Match 99.2%; Score 788; DB 22; Length 154;
Best Local Similarity 98.7%; Pred. No. 1.1e-76;
Matches 151; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 60
Db 2 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 61
QY 61 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
Db 62 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 121
QY 121 GDFGADAQAGMANKALELFRKDIAAAYKELGYQG 153
Db 122 GNFGADAQAGMANKALELFRKDIAAAYKELGYQG 154

RESULT 8

US-09-791-537-23088
; Sequence 23088, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23088
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb IMLR
US-09-791-537-23088

Query Match 99.0%; Score 786; DB 22; Length 154;
Best Local Similarity 98.7%; Pred. No. 1.8e-76;
Matches 151; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 60
Db 2 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 61
QY 61 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
Db 62 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 121
QY 121 GDFGADAQAGMANKALELFRKDIAAAYKELGYQG 153
Db 122 GNFGADAQAGMANKALELFRKDIAAAYKELGYQG 154

RESULT 9

US-09-791-537-23776
; Sequence 23776, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23776
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb 2SP0
US-09-791-537-23776

```

US-09-791-537-22333
Query Match 98.9%; Score 785; DB 22; Length 154;
Best Local Similarity 98.7%; Pred. No. 2.3e-76;
Matches 151; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLSGEQWLVLHWKAEADVAGHGQDILIRLPKSHPETLEKDFRPFKHLKTEAEMKASED 60
DB 2 VLSGEQWLVLHWKAEADVAGHGQDILIRLPKSHPETLEKDFRPFKHLKTEAEMKASED 61
QY 61 LKHGVTVLTAIGALLKKKGHHEARLKLPAOSHATKKKIPIKYLEFISEAIIHVLHSRHP 120
DB 62 LKHGVTVLTAIGALLKKKGHHEARLKLPAOSHATKKKIPIKYLEFISEAIIHVLHSRHP 121
QY 121 GDFGADAGAMNKALELFRKDIAAAYKELGYQG 153
DB 122 GNFADAGAMNKALELFRKDIAAAYKELGYQG 154

RESULT 12
US-09-791-537-23775
; Sequence 23775, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAN
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791.537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23775
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb 2SPL
US-09-791-537-23775

```

```

US-09-791-537-2375
Query Match          98.9%; Score 785; DB 22; Length 154;
Best Local Similarity 98.7%; Pred. No. 2.3e-76;
Matches 151; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 VLSGEQWLVLHWAKVEADVAGHGQDILRLFKSHPETLEKDFPFKHLKTEAEAKMSD 60
Db      2 VLSGEQWLVLHWAKVEADVAGHGQDILRLFKSHPETLEKDFPFKHLKTEAEAKMSD 61

QY      61 LKKGVTVLTALGAILKKKGHEAELKPLAQSHATKKKIPIKYLEFISEAIIHVLSRHP 120
Db      62 LKKGVTVLTALGAILKKKGHEAELKPLAQSHATKKKIPIKYLEFISEAIIHVLSRHP 121

QY      121 GFCGADAGGAMKALELFRKDIAAKYKELGYQG 153
Db      122 GFCGADAGGAMKALELFRKDIAAKYKELGYQG 154

RESULT 13
US-09-791-537-102151
; Sequence 102151, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAN
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 102151
; LENGTH: 154
; TYPE: PRT

```

```
; ORGANISM: pdb 1CH2A
US-09-791-537-102151

Query Match      98.9%; Score 785; DB 22; Length 154;
Best Local Similarity 98.7%; Pred. No. 2.3e-76;
Matches 151; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASD 60
Db 2 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASD 61
QY 61 LKKHGVTVLTALGAILKKKGHEAEELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
Db 62 LKKHGVTVLTALGAILKKKGHEAEELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 121
QY 121 GDFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
Db 122 GNFGADAQAGAMNKALELFRKDIAAKYKELGYQG 154

RESULT 14
US-09-791-537-102898
; Sequence 102898, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 102898
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb 1CP5A
US-09-791-537-102898

Query Match      98.9%; Score 785; DB 22; Length 154;
Best Local Similarity 98.7%; Pred. No. 2.3e-76;
Matches 151; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASD 60
Db 2 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASD 61
QY 61 LKKHGVTVLTALGAILKKKGHEAEELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
Db 62 LKKHGVTVLTALGAILKKKGHEAEELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 121
QY 121 GDFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
Db 122 GNFGADAQAGAMNKALELFRKDIAAKYKELGYQG 154

RESULT 15
US-09-791-537-51498
; Sequence 51498, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 51498
; LENGTH: 153
```

```
; TYPE: PRT
; ORGANISM: pdb 1DTWA
US-09-791-537-51498

Query Match      98.7%; Score 784; DB 22; Length 153;
Best Local Similarity 99.3%; Pred. No. 3e-76;
Matches 152; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASD 60
Db 1 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASD 60
QY 61 LKKHGVTVLTALGAILKKKGHEAEELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
Db 61 LKKHGVTVLTALGAILKKKGHEAEELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
QY 121 GDFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
Db 121 GDFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
```

Search completed: August 10, 2004, 15:39:00
Job time : 95.2821 secs

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:26:56 ; Search time 12.04 Seconds
(without alignments)
1368.899 Million cell updates/sec

Title: US-09-455-978b-76
Perfect score: 794
Sequence: 1 VLSEGEQLVLHVAKVEAD.....ALELFRKDIAAKYKELGYQG 153

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 618821

Minimum DB seq length: 0

Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW COMB.pcp:*
2: /cgn2_6/ptodata/2/paa/US06 NEW COMB.pcp:*
3: /cgn2_6/ptodata/2/paa/US07 NEW COMB.pcp:*
4: /cgn2_6/ptodata/2/paa/US08 NEW COMB.pcp:*
5: /cgn2_6/ptodata/2/paa/US09 NEW COMB.pcp:*
6: /cgn2_6/ptodata/2/paa/US10 NEW COMB.pcp:*
7: /cgn2_6/ptodata/2/paa/US60 NEW COMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	699	88.0	153	6	US-10-784-004-632
2	699	88.0	153	6	US-10-784-004-1039
3	699	88.0	154	6	US-10-170-205E-32946
4	699	88.0	154	6	US-10-170-205E-3330
5	699	88.0	154	6	US-10-170-205E-33331
6	647	81.5	154	6	US-10-784-004-310
7	647	81.5	154	6	US-10-784-004-898
8	472	59.4	113	6	US-10-793-479-4103
9	327	41.2	79	6	US-10-425-115-279565
10	314	39.5	71	6	US-10-793-479-4102
11	212.5	26.8	190	1	PCT-US04-07412-750
12	212.5	26.8	190	1	PCT-US02-17443-214
13	212.5	26.8	190	1	PCT-US02-17443-216
14	212.5	26.8	190	6	US-10-389-559-750
15	212.5	26.8	190	6	US-10-170-205E-19233
16	159	20.0	142	6	US-10-170-205E-7507
17	140	17.6	141	6	US-10-776-172-21
18	140	17.6	141	6	US-10-897-005-28
19	128	16.1	141	6	US-10-776-172-19
20	128	16.1	141	6	US-10-897-005-26
21	127	16.0	146	6	US-10-776-172-22
22	127	16.0	146	6	US-10-897-005-29
23	125	15.7	147	6	US-10-170-205E-23524
24	125	15.7	147	7	US-60-563-440-970
25	124	15.6	120	6	US-10-425-115-205641
26	124	15.6	147	1	PCT-US03-36002-201

ALIGNMENTS

RESULT 1

US-10-784-004-632
; Sequence 632, Application US/10784004
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 632
; LENGTH: 153
; TYPE: PRT
; ORGANISM: human
US-10-784-004-632
Query Match 88.0%; Score 699; DB 6; Length 153;
Best Local Similarity 84.9%; Pred No. 4.7e-64;
Matches 129; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY	2	LSEGEQLVLHVAKVEADVAGHQDILIRLFKSHPETLEKFDREKFKHLKTEAMKASEDL	61
Db	2	LSDGEQLVLNVWGKVEADIPGHQGVILIRLFKSHPETLEKFDREKFKHLKSEDEMKASEDL	61
QY	62	KHGVTVLTALGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG	121
Db	62	KHGVTVLTALGGILKKKGHEAEIKPLAQSHATKHKIPKYLEFISECIIQVLSKHPG	121
QY	122	DFGADAQGMNKALELFRKDIAAKYKELGYQG	153
Db	122	DFGADAQGMNKALELFRKDMASNYKELGFGQ	153

RESULT 2

US-10-784-004-1039
; Sequence 1039, Application US/10784004
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1039
; LENGTH: 153
; TYPE: PRT
; ORGANISM: human

US-10-784-004-1039

Query Match 88.0%; Score 699; DB 6; Length 153;
Best Local Similarity 84.9%; Pred. No. 4.7e-64;
Matches 129; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLPKSHPETLEKDFRPHKLTAEAMKASEDL 61
DB 2 LSDGEWQLVNVWGKVEADIPGHQGEVLIRLPKSHPETLEKDFRPHKLTAEAMKASEDL 61
QY 62 KHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSHRHPG 121
DB 62 KHGATVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSHRHPG 121

QY 122 DFGADAQAGAMNKALELFRKDIKAAKYKELGYQG 153
DB 122 DFGADAQAGAMNKALELFRKDMASNYKELGFQG 153

RESULT 3

US-10-170-205E-32946
; Sequence 32946, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 32946
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-32946

Query Match 88.0%; Score 699; DB 6; Length 154;
Best Local Similarity 84.9%; Pred. No. 4.7e-64;
Matches 129; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLPKSHPETLEKDFRPHKLTAEAMKASEDL 61
DB 3 LSDGEWQLVNVWGKVEADIPGHQGEVLIRLPKSHPETLEKDFRPHKLTAEAMKASEDL 62
QY 62 KHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSHRHPG 121
DB 63 KHGATVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSHRHPG 122

QY 122 DFGADAQAGAMNKALELFRKDIKAAKYKELGYQG 153
DB 123 DFGADAQAGAMNKALELFRKDMASNYKELGFQG 154

RESULT 4

US-10-170-205E-33330
; Sequence 33330, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 33330
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-33330

Query Match 88.0%; Score 699; DB 6; Length 154;

Best Local Similarity 84.9%; Pred. No. 4.7e-64;
Matches 129; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLPKSHPETLEKDFRPHKLTAEAMKASEDL 61
DB 3 LSDGEWQLVNVWGKVEADIPGHQGEVLIRLPKSHPETLEKDFRPHKLTAEAMKASEDL 62
QY 62 KHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSHRHPG 121
DB 63 KHGATVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSHRHPG 122
QY 122 DFGADAQAGAMNKALELFRKDIKAAKYKELGYQG 153
DB 123 DFGADAQAGAMNKALELFRKDMASNYKELGFQG 154

RESULT 5

US-10-170-205E-33331
; Sequence 33331, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 33331
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-33331

Query Match 88.0%; Score 699; DB 6; Length 154;
Best Local Similarity 84.9%; Pred. No. 4.7e-64;
Matches 129; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLPKSHPETLEKDFRPHKLTAEAMKASEDL 61
DB 3 LSDGEWQLVNVWGKVEADIPGHQGEVLIRLPKSHPETLEKDFRPHKLTAEAMKASEDL 62
QY 62 KHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSHRHPG 121
DB 63 KHGATVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSHRHPG 122
QY 122 DFGADAQAGAMNKALELFRKDIKAAKYKELGYQG 153
DB 123 DFGADAQAGAMNKALELFRKDMASNYKELGFQG 154

RESULT 6

US-10-784-004-310
; Sequence 310, Application US/10784004
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 310
; LENGTH: 154
; TYPE: PRT
; ORGANISM: rat
US-10-784-004-310

Query Match 81.5%; Score 647; DB 6; Length 154;
Best Local Similarity 78.9%; Pred. No. 1e-58;
Matches 120; Conservative 18; Mismatches 14; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLPKSHPETLEKDFRPHKLTAEAMKASEDL 61

Db 3 LSDGEWQVNLINWGVKVEADVAGHGDILIRLFSKHPETLEKFDKPKNLKSEEMKASDDL 62
QY 62 KKHGVTVTALGAILKKKGHHEAEIKPLAQSHATKHKIPKYLEFISEAIHHVLSRHPG 121
Db 63 KKHGCTVTALTGILTKKGQHAETIQPLAQSHATKHKIPVKYLEFISEVILQVLKKRYSG 122

QY 122 DFGADAQAGMKALELFRKDIAAKYKELGYQG 153
Db 123 DFGADAQAGMKALELFRNDIAAKYKELGFQG 154

RESULT 7
US-10-784-004-898
; Sequence 898, Application US/10784004
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 898
; LENGTH: 154
; TYPE: PRT
; ORGANISM: rat
US-10-784-004-898

Query Match 81.5%; Score 647; DB 6; Length 154;
Best Local Similarity 78.9%; Pred. No. 1e-58;
Matches 120; Conservative 18; Mismatches 14; Indels 0; Gaps 0;

QY 2 LSEGEWQVNLINWGVKVEADVAGHGDILIRLFSKHPETLEKFDKPKHLKTEEMKASDDL 61
Db 3 LSDGEWQVNLINWGVKVEADVAGHGDILIRLFSKHPETLEKFDKPKNLKSEEMKASDDL 62

QY 62 KKHGVTVTALGAILKKKGHHEAEIKPLAQSHATKHKIPKYLEFISEAIHHVLSRHPG 121
Db 63 KKHGCTVTALTGILTKKGQHAETIQPLAQSHATKHKIPVKYLEFISEVILQVLKKRYSG 122

QY 122 DFGADAQAGMKALELFRKDIAAKYKELGYQG 153
Db 123 DFGADAQAGMKALELFRNDIAAKYKELGFQG 154

RESULT 8
US-10-793-479-4103
; Sequence 4103, Application US/10793479
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/10/793,479
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: US/09/513,999
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 4103
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -20...-1
; OTHER INFORMATION: score 3.5
; OTHER INFORMATION: seq WQVLNWMVGKVEA/DI
; FEATURE:

; NAME/KEY: UNSURE
; LOCATION: 84
; OTHER INFORMATION: Xaa-Phe or Ile or Leu or Met or Val
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 85
; OTHER INFORMATION: Xaa-Pro or Thr
US-10-793-479-4103

Query Match 59.4%; Score 472; DB 6; Length 113;
Best Local Similarity 85.1%; Pred. No. 6.3e-41;
Matches 86; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

QY 2 LSEGEWQVNLINWGVKVEADVAGHGDILIRLFSKHPETLEKFDKPKHLKTEEMKASDDL 61
Db 3 LSDGEWQVNLINWGVKVEADIPGHGQEVILIRLFPKHPETLEKFDKPKHLKSEDEMKASDDL 62

QY 62 KKHGVTVTALGAILKKKGHHEAEIKPLAQSHATKHKIPK 102
Db 63 KKHGATVTALTGILTKKGHHEAEIKPLAQSHATKHKIPVK 103

RESULT 9
US-10-425-115-279565
; Sequence 279565, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 279565
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_106543C.1.pep
US-10-425-115-279565

Query Match 41.2%; Score 327; DB 6; Length 79;
Best Local Similarity 76.6%; Pred. No. 3e-26;
Matches 59; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 2 LSEGEWQVNLINWGVKVEADVAGHGDILIRLFSKHPETLEKFDKPKHLKTEEMKASDDL 61
Db 3 LSDGEWQVNLINWGVKVEADLGHGQEVILISLFKAHPETLEKFDKPKNLKSEEMKSSDDL 62

QY 62 KKHGVTVTALGAILKK 78
Db 63 KKHGCTVTALTGILTKK 79

RESULT 10
US-10-793-479-4102
; Sequence 4102, Application US/10793479
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/10/793,479
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: US/09/513,999
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681

```
; SOFTWARE: Patent.pm
; SEQ ID NO 4102
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -20...-1
; OTHER INFORMATION: score 3.5
; OTHER INFORMATION: seq WOLVLNVWGKVEA/DI
US-10-793-479-4102

Query Match          39.5%; Score 314; DB 6; Length 71;
Best Local Similarity 82.6%; Pred. No. 5.7e-25;
Matches 57; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKFHLKTEAEKASEDL 61
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 LSDGEWQLVNVWGKVEADPGHGQEVILIRLFKSHPETLEKDFKPKHLKSEDEKASEDL 62
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 KKHGVTVLT 70
   |||||:|||||
Db 63 KKHGATVLT 71
   |||||:|||||

RESULT 11
PCT-US04-07412-750
; Sequence 750, Application PC/TUS0407412
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-rui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Ghosh, Malabika
; APPLICANT: Xue, Aidong J.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunrui
; APPLICANT: Goodrich, Ryle W.
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wehrman, Tom
; APPLICANT: Weng, Gezhi
; APPLICANT: Wang, Zhiwei
; APPLICANT: Boyle, Bryan J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 822CIP/PCT
; CURRENT APPLICATION NUMBER: PCT/US04/07412
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: US 10/389,559
; PRIOR FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 60/365,264
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/340,187
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 10/296,115
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 10/275,027
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1920

; SOFTWARE: pt_FL_genes Version 6.0
; SEQ ID NO 750
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Homo sapiens
; PCT-US04-07412-750

Query Match          26.8%; Score 212.5; DB 1; Length 190;
Best Local Similarity 31.6%; Pred. No. 5.4e-14;
Matches 49; Conservative 31; Mismatches 68; Indels 7; Gaps 2;

QY 2 LSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKFHLKTEAEKASEDL 61
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 19 LSEAEKAVQAWALYANCEDVGVALIVRFVNFPSAKQYFSQFKHMDPLEMERSPOL 78
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 KKHGVTVLTALGAILKKKHHEAE-----LKPASHATKHKIPIKYLEFISEAIHVLH 116
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 79 RKHACRVMGALNTVVENL--HDPDKVSSVLAIVGKAHALKHKVEPVYFKILSGVILEVVA 136
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 117 SRHPGDFGADAOGAMNKALELFRKDIAAKYKELGY 151
   |||||:|||||
Db 137 EEFASDFPPETORAWAKLRGLIYSHVTAAYKEVGW 171
   |||||:|||||

RESULT 12
PCT-US02-17443-214
; Sequence 214, Application PC/TUS0217443
; GENERAL INFORMATION:
; APPLICANT: CuraGen Corporation
; APPLICANT: Agee, Michele
; APPLICANT: Alsobrook, John
; APPLICANT: Anderson, David
; APPLICANT: Berghs, Constance
; APPLICANT: Burgess, Catherine
; APPLICANT: Casman, Stacie
; APPLICANT: Catterton, Elina
; APPLICANT: DiPippo, Vincent
; APPLICANT: Edinger, Shlomit
; APPLICANT: Eisen, Andrew
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Herrmann, John
; APPLICANT: Hjalt, Tord
; APPLICANT: Ji, Weizhen
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Khrantsov, Nikolai
; APPLICANT: Lepley, Denise
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Malyankar, Uriel
; APPLICANT: Miller, Charles
; APPLICANT: Millet, Isabelle
; APPLICANT: Ort, Tatiana
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Pena, Carol
; APPLICANT: Rastelli, Luca
; APPLICANT: Rieger, Daniel
; APPLICANT: Rothenberg, Mark
; APPLICANT: Shenoy, Suresh
; APPLICANT: Shinkets, Richard
; APPLICANT: Smithson, Glennda
; APPLICANT: Spaderina, Steven
; APPLICANT: Spytek, Kimberly
; APPLICANT: Stone, David
; APPLICANT: Vernet, Corine
; APPLICANT: Zhong, Haihong
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-377C-061
```

```

; APPLICANT: Lepley, Denise
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Malvankar, Uriel
; APPLICANT: Miller, Charles
; APPLICANT: Millet, Isabelle
; APPLICANT: Ort, Tatiana
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Pena, Carol
; APPLICANT: Rastelli, Luca
; APPLICANT: Rieger, Daniel
; APPLICANT: Rothenberg, Mark
; APPLICANT: Shenoy, Suresh
; APPLICANT: Shimkets, Richard
; APPLICANT: Smithson, Glenna
; APPLICANT: Spaderna, Steven
; APPLICANT: Spytek, Kimberly
; APPLICANT: Stone, David
; APPLICANT: Vernet, Corine
; APPLICANT: Zhong, Haihong
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-377C-061
; CURRENT APPLICATION NUMBER: PCT/US02/17443
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: 10/160,619
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: 60/295,661
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/295,607
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/296,404
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/296,418
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/296,575
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/297,414
; PRIOR FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: 60/297,573
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/297,567
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/298,285
; PRIOR FILING DATE: 2001-06-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 539
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 216
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-17443-216

Query Match          26.8%; Score 212.5; DB 1; Length 190;
Best Local Similarity 31.6%; Pred. No. 5.4e-14; Indels 7; Gaps 2;
Matches 49; Conservative 31; Mismatches 68;

Qy 2 LSEGEWLVLHVWAKVEADYAGHGQDILIRLKSHPETLEKFDPRFKHLKTEAMKASEDL 61
Db 19 LSEAEKAVQAMWRLYANVEDGVAILVRFVFNFPSSAKQVFSQFKHWDPLEMERSQL 78

Qy 62 KKGVTVTLTALGAILKKKGHEAE-----LKPLAQSHATKHKIPIKYLEFTSEAIHVLH 116
Db 79 RGHACRVMGALNTVENVL--HDPDKVSVLALVGKALKHKVDPVYFKILSGVILEVA 136

Qy 117 SRHPGDFGADAQGMNKALELFRKDIKAYKELGY 151
Db 137 EEFASDFPPPTQRAWAKLRGLIYSHVTAAYKEVGW 171

RESULT 14

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:22:45 ; Search time 6.66828 Seconds
(without alignments)
2207.061 Million cell updates/sec

Title: US-09-455-978b-76

Perfect score: 794

Sequence: 1 VLSEGEWQLVHLHWAKVEAD.....ALELPRKDIAAKYKELGYQG 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	794	100.0	153	1 MYWHP	myoglobin [validat
2	773	97.4	153	1 MYWHW	myoglobin - dwarf
3	746	94.0	153	1 MYWHC	myoglobin - Califo
4	740	93.2	153	1 MYWHH	myoglobin - humpha
5	738	92.9	153	1 MYDDAR	myoglobin - Amazon
6	737	92.8	153	1 MYWHT	myoglobin - pilot
7	734	92.4	153	1 MYWHL	myoglobin - killer
8	733	92.3	153	1 MYDD	myoglobin - bottle
9	733	92.3	153	1 MYDDBS	myoglobin - saddle
10	731	92.1	153	1 MYPE	myoglobin - harbor
11	731	92.1	153	1 MYPED	myoglobin - Dall's
12	731	92.1	153	1 MYWHK	myoglobin - minke
13	726	91.4	153	1 MYWHF	myoglobin - finbac
14	725	91.3	153	1 MYWHU	myoglobin - Hubbs'
15	722	90.9	153	1 MYWHZ	myoglobin - goose-
16	722	90.9	153	2 JN0410	myoglobin - Europe
17	712	89.7	153	1 MYHO	myoglobin [validat
18	712	89.7	153	1 MYHOZ	myoglobin - common
19	707	89.0	153	1 MYCZ	myoglobin - chimpa
20	705	88.8	153	1 MYLP	myoglobin - potto
21	703	88.5	153	1 MYRB	myoglobin - rabbit
22	701	88.3	153	1 MYNO126	myoglobin - mounta
23	701	88.3	153	1 MYGO	myoglobin - muskra
24	700	88.2	153	1 MYPG	myoglobin [validat
25	699	88.0	154	1 MYHU	myoglobin [validat
26	695	87.5	153	1 MYQIM	myoglobin - crab-e
27	694	87.4	153	1 MYOG	myoglobin - Bornea
28	693	87.3	153	1 MYGI	myoglobin - agile
29	693	87.3	153	1 MYTS	myoglobin - common

ALIGNMENTS

RESULT 1

MYWHP

myoglobin [validated] - sperm whale

C:Species: Physeter catodon (sperm whale)

C:Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 15-Sep-2000

C:Accession: A90591; A92844; A93150; A02506

R:Romero-Herrera, A.E.; Lehmann, H.

Biochim. Biophys. Acta 336, 318-323, 1974

A:Title: Residue 122 of sperm whale and horse myoglobin.

A:Reference number: A90591

A:Accession: A90591

A:Molecule type: protein

A:Residues: 1-153 <ROM>

A:Experimental source: skeletal muscle

R:Takano, T.

J. Mol. Biol. 110, 537-568, 1977

A:Title: Structure of myoglobin refined at 2.0 angstrom resolution. I. Crystallographic

A:Reference number: A92844; MUID:7714097; PMID:845959

A:Contents: X-ray crystallography of metmyoglobin, 2.0 angstroms

A:Accession: A92844

A:Molecule type: protein

A:Residues: 1-120,'A',122-153 <TAK>

R:Edmundson, A.B.

Nature 205, 883-887, 1965

A:Title: Amino-acid sequence of sperm whale myoglobin.

A:Reference number: A93150

A:Accession: A93150

A:Molecule type: protein

A:Residues: 1-121,'N',123-153 <EDM>

A:Experimental source: heart muscle

R:Watson, H.C.; Kendrew, J.C.

submitted to the Brookhaven Protein Data Bank, April 1973

A:Reference number: A50272; PDB:1MEN

A:Contents: annotation; X-ray crystallography, 2.0 angstroms, residues 1-153

R:Takano, T.

submitted to the Brookhaven Protein Data Bank, January 1988

A:Reference number: A50732; PDB:5MEN

A:Contents: annotation; X-ray crystallography, 2.0 angstroms, residues 1-153

J. Mol. Biol. 110, 569-584, 1977

A:Title: Structure of myoglobin refined at 2.0 angstrom resolution. II. Structure of deo

A:Reference number: A92845; MUID:7714098; PMID:845960

A:Contents: annotation; X-ray crystallography of deoxymyoglobin, 2.0 angstroms

C:Function:

A:Description: binds molecular oxygen for intracellular storage and transport, primarily

C:Superfamily: globin; globin homology

C:Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier

F:2-147/Domain: Globin homology <GLB>

F:64/Binding site: oxygen (His) (distal axial ligand) #status experimental

F:93/Binding site: heme iron (His) (proximal axial ligand) #status experimental

Query Match 100.0%; Score 794; DB 1; Length 153;

myoglobin - siaman
myoglobin - olive
myoglobin - red gu
myoglobin - hanuma
myoglobin - dourou
myoglobin - common
myoglobin - Egypti
myoglobin - Eurasi
myoglobin - weasel
myoglobin - aardva
myoglobin - slow l
myoglobin - Califo
myoglobin - northe
myoglobin - gray s
myoglobin [validat
myoglobin - thick-

30 693 87.3 153 2 A61366
31 691 87.0 153 1 MYBAO
32 691 87.0 153 1 MYQGR
33 691 87.0 153 1 MYMQL
34 691 87.0 153 1 MYMQN
35 690 86.9 153 1 MYCJ
36 690 86.9 153 1 MYBTF
37 688 86.6 153 1 MYOT
38 688 86.6 153 1 MYLEM
39 687 86.5 153 1 MYOY
40 685 86.3 153 1 MYLR
41 684 86.1 153 1 MYZC
42 683 86.0 153 1 MYETNG
43 682 85.9 153 1 MYSLG
44 682 85.9 153 1 MYSLH
45 679 85.5 153 1 MYGC

Best Local Similarity 100.0%; Pred. No. 8.4e-59;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60
DB 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60

QY 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
DB 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120

QY 121 GDFGADAQAMNKALELFRKDTAAKYKELGYQG 153
DB 121 GDFGADAQAMNKALELFRKDTAAKYKELGYQG 153

RESULT 2

MYWHH

myoglobin - dwarf sperm whale
C:Species: Kogia simus (dwarf sperm whale)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 03-Mar-2000
C:Accession: A90407; A02506
R:Lehman, L.D.; Jones, B.N.; Lehman, L.D.; Gurd, F.R.N.
Biochemistry 16, 873-877, 1977
A:Title: The complete amino acid sequence of the major component myoglobin of dwarf sperm whale.
A:Reference number: A90407; MUID:77134684; PMID:843520
A:Accession: A90407
A:Molecule type: protein
A:Residues: 1-153 <DMU>
C:Superfamily: globin; globin homology
C:Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F:2-147/Domain: globin homology <GLB>
F:64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F:93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 97.4%; Score 773; DB 1; Length 153;
Best Local Similarity 96.7%; Pred. No. 4.6e-57;
Matches 148; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60
DB 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60

QY 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
DB 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120

QY 121 GDFGADAQAMNKALELFRKDTAAKYKELGYQG 153
DB 121 ADFGADAQAMNKALELFRKDTAAKYKELGYQG 153

RESULT 3

MYWHC

myoglobin - California gray whale
C:Species: Eschrichtius robustus, Eschrichtius gibbosus (California gray whale)
C:Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 03-Mar-2000
C:Accession: A02499
R:Bogardt Jr., R.A.; Dwulet, F.E.; Jones, B.N.; Lehman, L.D.; Gurd, F.R.N.
Biochemistry 15, 2597-2602, 1976
A:Title: Complete primary structure of the major component myoglobin of California gray whale.
A:Reference number: A02499; MUID:76232191; PMID:938629
A:Accession: A02499
A:Molecule type: protein
A:Residues: 1-153 <BOG>
C:Comment: This myoglobin was isolated from skeletal muscle.
C:Superfamily: globin; globin homology
C:Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F:2-147/Domain: globin homology <GLB>
F:64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F:93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 94.0%; Score 746; DB 1; Length 153;

Best Local Similarity 92.8%; Pred. No. 7.8e-55;
Matches 142; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60
DB 1 VLSDAEWQLVLNIWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60

QY 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
DB 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120

QY 121 GDFGADAQAMNKALELFRKDTAAKYKELGYQG 153
DB 121 GDFGADAQAMNKALELFRKDTAAKYKELGYQG 153

RESULT 4

MYWHH

myoglobin - humpback whale
C:Species: Megaptera novaeangliae (humpback whale)
C:Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
C:Accession: A90416; A91088; A02500
R:Lehman, L.D.; Dwulet, F.E.; Jones, B.N.; Bogardt Jr., R.A.; Krueckeberg, S.T.; Vissache
Biochemistry 17, 3736-3739, 1978
A:Title: Complete amino acid sequence of the major component myoglobin from the humpback whale.
A:Reference number: A90416; MUID:79021601; PMID:698193
A:Accession: A90416
A:Molecule type: protein
A:Residues: 1-153 <LEH>
R:Edman, P.; Beggs, G.

Eur. J. Biochem. 1, 80-91, 1967
A:Title: A protein sequenator.
A:Reference number: A91088; MUID:68049168; PMID:6059350
A:Accession: A91088
A:Molecule type: protein
A:Residues: 1-60 <EDM>

A:Note: This was the first sequence determined using the modern sequenator
C:Comment: This myoglobin was isolated from skeletal muscle.
C:Superfamily: globin; globin homology
C:Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F:2-147/Domain: globin homology <GLB>

F:64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F:93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 93.2%; Score 740; DB 1; Length 153;
Best Local Similarity 92.2%; Pred. No. 2.5e-54;
Matches 141; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60
DB 1 VLSDAEWQLVLNIWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60

QY 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
DB 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120

QY 121 GDFGADAQAMNKALELFRKDTAAKYKELGYQG 153
DB 121 ADFGADAQAMNKALELFRKDTAAKYKELGYQG 153

RESULT 5

MYDDAR

myoglobin - Amazon dolphin
C:Species: Inia geoffrensis (Amazon dolphin)
C:Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
C:Accession: A02503
R:Dwulet, F.E.; Bogardt, R.A.; Jones, B.N.; Lehman, L.D.; Gurd, F.R.N.
Biochemistry 14, 5336-5343, 1975
A:Title: The complete amino acid sequence of the major component myoglobin of Amazon river dolphin.
A:Reference number: A02503; MUID:76062412; PMID:1191640
A:Accession: A02503
A:Molecule type: protein

A;Residues: 1-153 <DWU>
 C;Comment: This myoglobin was isolated from skeletal muscle.
 C;Superfamily: globin; globin homology
 C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
 F;2-147/Domain: globin homology <GLB>
 F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
 F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 92.9%; Score 738; DB 1; Length 153;
 Best Local Similarity 91.4%; Pred. No. 3.6e-54;
 Matches 139; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 2 LSEGEQVLVHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKASDDL 61
 DB 2 LSDGEQVLVNWGKVEADLAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKASDDL 61

QY 62 KKHGVTVTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG 121
 DB 62 KKHGNTVTALGGILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG 121

QY 122 DFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
 DB 122 DFGADAQAGAMNKALELFRKDIAAKYKELGFHG 153

RESULT 6
 MYWHT
 myoglobin - pilot whale
 C;Species: Globicephala melaena (pilot whale)
 C;Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
 C;Accession: A02496
 R;Jones, B.N.; Dwulet, F.E.; Lehman, L.D.; Garner, M.H.; Bogardt Jr., R.A.; Garner, W.H.
 Biochemistry 17, 1971-1974, 1978
 A;Title: Complete amino acid sequence of myoglobin from the pilot whale, Globicephala melaena
 A;Reference number: A02496; PMID:78187213; PMID:656376

A;Molecule type: protein
 A;Residues: 1-153 <JON>
 C;Comment: This myoglobin was isolated from skeletal muscle.
 C;Superfamily: globin; globin homology
 C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
 F;2-147/Domain: globin homology <GLB>
 F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
 F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 92.8%; Score 737; DB 1; Length 153;
 Best Local Similarity 92.1%; Pred. No. 4.3e-54;
 Matches 140; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 LSEGEQVLVHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKASDDL 61
 DB 2 LSDGEQVLVNWGKVEADLAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKASDDL 61

QY 62 KKHGVTVTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG 121
 DB 62 KKHGNTVTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPA 121

QY 122 DFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
 DB 122 DFGADAQAGAMNKALELFRKDIAAKYKELGFHG 153

RESULT 7
 MYWHL
 myoglobin - killer whale
 C;Species: Orcinus orca (killer whale)
 C;Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
 C;Accession: A92956; A90614; A02495
 J;Meuth, J.L.; Jones, B.N.; Gurd, F.R.N.
 J. Mol. Evol. 17, 163-166, 1981
 A;Title: Reassignment of residue 122 in the myoglobin from the killer whale, Orcinus orca
 A;Reference number: A92956; PMID:81267443; PMID:6115067
 A;Accession: A92956

A;Molecule type: protein
 A;Residues: 1-153 <MEU>
 R;Castillo, O.; Lehmann, H.; Jones, L.T.
 Biochim. Biophys. Acta 491, 23-28, 1977
 A;Title: The myoglobin of the killer whale (Orcinus orca).
 A;Reference number: A90614; PMID:77134898; PMID:849459
 A;Accession: A90614

A;Molecule type: protein
 A;Residues: 1-121, Q;123-153 <CAS>
 C;Comment: This myoglobin was isolated from skeletal muscle.
 C;Superfamily: globin; globin homology
 C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
 F;2-147/Domain: globin homology <GLB>
 F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
 F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 92.4%; Score 734; DB 1; Length 153;
 Best Local Similarity 91.4%; Pred. No. 7.7e-54;
 Matches 139; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 2 LSEGEQVLVHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKASDDL 61
 DB 2 LSDGEQVLVNWGKVEADLAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKASDDL 61

QY 62 KKHGVTVTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG 121
 DB 62 KKHGNTVTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPA 121

QY 122 DFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
 DB 122 DFGADAQAGAMNKALELFRKDIAAKYKELGFHG 153

RESULT 8
 MYDD
 myoglobin - bottle-nosed dolphin
 C;Species: Tursiops truncatus (bottle-nosed dolphin)
 C;Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 03-Mar-2000
 C;Accession: A02494
 R;Jones, B.N.; Vigna, R.A.; Dwulet, F.E.; Bogardt, R.A.; Lehman, L.D.; Gurd, F.R.N.
 Biochemistry 15, 4418-4422, 1976
 A;Title: Complete amino acid sequence of the myoglobin from the Atlantic bottlenosed dolphin
 A;Reference number: A02494; PMID:77022063; PMID:974068
 A;Accession: A02494

A;Molecule type: protein
 A;Residues: 1-153 <JON>
 C;Superfamily: globin; globin homology
 C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
 F;2-147/Domain: globin homology <GLB>
 F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
 F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 92.3%; Score 733; DB 1; Length 153;
 Best Local Similarity 90.8%; Pred. No. 9.3e-54;
 Matches 138; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

QY 2 LSEGEQVLVHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKASDDL 61
 DB 2 LSDGEQVLVNWGKVEADLAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKASDDL 61

QY 62 KKHGVTVTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG 121
 DB 62 KKHGNTVTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPA 121

QY 122 DFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
 DB 122 DFGADAQAGAMNKALELFRKDIAAKYKELGFHG 153

RESULT 9
 MYDDBS
 myoglobin - saddleback dolphin
 C;Species: Delphinus delphis (saddleback dolphin)

A>Note: the saddleback dolphin is also called Black Sea dolphin
C>Date: 27-Nov-1985 #sequence revision 31-Mar-1991 #text_change 03-Mar-2000
C/Accession: A26230; A02497; A37509; A37510; A02494
R/Wang, C.C.; Avila, R.; Jones, B.N.; Gurd, F.R.N.
Biochemistry 15, 4978-4981, 1977
A>Title: Complete primary structure of the major component myoglobin of Pacific common dolphin
A/Reference number: A26230; MUID:178020893; PMID:911808
A/Accession: A26230
A/Molecule type: protein
A/Residues: 1-153 <WAN>
R/Kluh, I.; Bakardjieva, A.
FEBS Lett. 17, 31-34, 1971
A>Title: Primary structure of N-terminal part of molecule of dolphin myoglobin.
A/Reference number: A02497
A/Accession: A02497
A/Molecule type: protein
A/Residues: 1-20,'V',22-27,'I',29-31 <KLU>
R/Bakardjieva, M.; Nedkov, P.; Bakardjieva, A.; Genov, N.
Biochim Biophys Acta 221, 136-139, 1970
A>Title: Difference in amino acid sequence between dolphin and sperm whale myoglobins.
A/Reference number: A37509; MUID:71014229; PMID:5473803
A/Accession: A37509
A/Molecule type: protein
A/Residues: 31-'65','D',67-121,'Q',123-153 <KAR>
A>Note: The amidation states of residues 60, 85, and 132 were not determined and were as submitted to the Atlas, December 1977
R/Kluh, I.
A/Reference number: A37510
A/Contents: revision to residue 26
A/Accession: A37510
A/Molecule type: protein
A/Residues: 26 <KL2>
C/Superfamily: globin; globin homology
C/Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F:2-147/Domain: globin homology <GLB>
F:64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F:93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 92.3%; Score 733; DB 1; Length 153;
Best Local Similarity 90.8%; Pred. No. 9.3e-54;
Matches 138; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

Qy 2 LSEGEQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFPKHLKTEAMKASDL 61
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 2 LSDGEQLVLNVWGKVADLADHGQDVLIIRLFKGHPETLEKDFPKHLKTEADMKASDL 61
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Qy 62 KHGCVTVLTALGAILKKKHHEELPLAQASHATKHPIKYLEFISEAIHVLHSRPG 121
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 62 KHGNVTLVLTALGAILKKKHGHDELPLAQASHATKHPIKYLEFISEAIHVLHSRPA 121
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Qy 122 DFGADAQGAMNKALELFRKDIAAKYKELGYQG 153
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 122 EFGADAQGAMNKALELFRKDIAAKYKELGFHG 153
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 10
MYPE
myoglobin - harbor porpoise
C/Species: Phocoena phocoena (harbor porpoise)
C/Date: 27-Nov-1985 #sequence revision 27-Nov-1985 #text_change 03-Mar-2000
C/Accession: B92045; A90415; A02498
R/Bradshaw, R.A.; Gurd, F.R.N.
J. Biol. Chem. 244, 2167-2181, 1969
A>Title: Comparison of myoglobins from harbor seal, porpoise, and sperm whale.
A/Reference number: A92045; MUID:69177451; PMID:5782005
A/Accession: B92045
A/Molecule type: protein
A/Residues: 1-82,'EAN',86-153 <BRA>
A>Note: This sequence has been revised in reference A90415
A/Note: This is the final paper in a series
R/Meuth, J.L.; Jones, B.N.; Garner, W.H.; Gurd, F.R.N.
Biochemistry 17, 3429-3431, 1978
A>Title: Complete amino acid sequence of the myoglobin from the Dall porpoise (Phocoenoides dalli)

A;Title: The complete amino acid sequence of the major component myoglobin from the Arch
A;Reference number: A02501; MUID:77112438; PMID:836810
A;Accession: A02501
A;Molecule type: protein
A;Residues: 1-153 <LEH>
C;Comment: This myoglobin was isolated from skeletal muscle.
C;Superfamily: globin; globin homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F;2-147/Domain: globin homology <GLB>
F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 92.1%; Score 731; DB 1; Length 153;
Best Local Similarity 90.8%; Pred. No. 1.4e-53;
Matches 139; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDPRFKHLKTEAEKASEDL 60
Db 1 VLSDAEWHVLVNIWAKVEADVAGHQDILIRLPKSHPTLEKFDPRFKHLKTEAEKASEDL 60
QY 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
Db 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
QY 121 GDFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153
Db 121 ADFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 13
MYWHF
myoglobin - finback whale
C;Species: Balaeoptera physalus (finback whale, common rorqual)
C;Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
C;Accession: A02502
R;DiMarchi, R.D.; Wang, C.C.; Hemenway, J.B.; Gurd, F.R.N.
Biochemistry 17, 1968-1970, 1978
A;Title: Complete amino acid sequence of the major component myoglobin of finback whale
A;Reference number: A02502; MUID:78187212; PMID:656375
A;Accession: A02502
A;Molecule type: protein
A;Residues: 1-153 <DIM>
C;Comment: This myoglobin was isolated from skeletal muscle.
C;Superfamily: globin; globin homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F;2-147/Domain: globin homology <GLB>
F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 91.4%; Score 726; DB 1; Length 153;
Best Local Similarity 90.2%; Pred. No. 3.5e-53;
Matches 138; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDPRFKHLKTEAEKASEDL 60
Db 1 VLSDAEWHVLVNIWAKVEADVAGHQDILIRLPKSHPTLEKFDPRFKHLKTEAEKASEDL 60
QY 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
Db 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
QY 121 GDFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153
Db 121 ADFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 14
MYWHU
myoglobin - Hubbs' whale
C;Species: Mesoplodon carlhubbsi (Hubbs' whale)
C;Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
C;Accession: A02505
R;Dwulet, J.A.; Dwulet, F.E.; Gurd, F.R.N.

A;Title: The complete amino acid sequence of the major component myoglobin from the Arch
A;Reference number: A02501; MUID:77112438; PMID:836810
A;Accession: A02501
A;Molecule type: protein
A;Residues: 1-153 <LEH>
C;Comment: This myoglobin was isolated from skeletal muscle.
C;Superfamily: globin; globin homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F;2-147/Domain: globin homology <GLB>
F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 92.1%; Score 731; DB 1; Length 153;
Best Local Similarity 90.8%; Pred. No. 1.4e-53;
Matches 139; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDPRFKHLKTEAEKASEDL 60
Db 1 VLSDAEWHVLVNIWAKVEADVAGHQDILIRLPKSHPTLEKFDPRFKHLKTEAEKASEDL 60
QY 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
Db 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
QY 121 GDFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153
Db 121 ADFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 13
MYWHF
myoglobin - finback whale
C;Species: Balaeoptera physalus (finback whale, common rorqual)
C;Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
C;Accession: A02502
R;DiMarchi, R.D.; Wang, C.C.; Hemenway, J.B.; Gurd, F.R.N.
Biochemistry 17, 1968-1970, 1978
A;Title: Complete amino acid sequence of the major component myoglobin of finback whale
A;Reference number: A02502; MUID:78187212; PMID:656375
A;Accession: A02502
A;Molecule type: protein
A;Residues: 1-153 <DIM>
C;Comment: This myoglobin was isolated from skeletal muscle.
C;Superfamily: globin; globin homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F;2-147/Domain: globin homology <GLB>
F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 91.4%; Score 726; DB 1; Length 153;
Best Local Similarity 90.2%; Pred. No. 3.5e-53;
Matches 138; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDPRFKHLKTEAEKASEDL 60
Db 1 VLSDAEWHVLVNIWAKVEADVAGHQDILIRLPKSHPTLEKFDPRFKHLKTEAEKASEDL 60
QY 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
Db 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
QY 121 GDFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153
Db 121 ADFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 14
MYWHU
myoglobin - Hubbs' whale
C;Species: Mesoplodon carlhubbsi (Hubbs' whale)
C;Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
C;Accession: A02505
R;Dwulet, J.A.; Dwulet, F.E.; Gurd, F.R.N.

Biochim. Biophys. Acta 624, 121-129, 1980
A;Title: Complete amino acid sequence of the major component myoglobin from Hubb's beake
A;Reference number: A02505; MUID:8100592; PMID:7407230
A;Accession: A02505
A;Molecule type: protein
A;Residues: 1-153 <DMU>
C;Comment: This myoglobin was isolated from skeletal muscle.
C;Superfamily: globin; globin homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F;2-147/Domain: globin homology <GLB>
F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 91.3%; Score 725; DB 1; Length 153;
Best Local Similarity 90.1%; Pred. No. 4.3e-53;
Matches 137; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDPRFKHLKTEAEKASEDL 61
Db 2 LSEAEWQLVHVWAKVEADLSGHGOEILIRLPKSHPTLEKFDPRFKHLKTEAEKASEDL 61
QY 62 KKHGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG 121
Db 62 KKHGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG 121
QY 122 DFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153
Db 122 DFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 15
MYWHZ
myoglobin - goose-beaked whale
C;Species: Ziphius cavirostris (goose-beaked whale)
C;Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 03-Mar-2000
C;Accession: A02504
R;Lehman, L.D.; Jones, B.N.; Dwulet, F.E.; Bogardt Jr., R.A.; Gurd, F.R.N.
Biochim. Biophys. Acta 625, 221-229, 1980
A;Title: Complete amino acid sequence of the major component myoglobin from the goose-be
A;Reference number: A02504; MUID:81063290; PMID:7437458
A;Accession: A02504
A;Molecule type: protein
A;Residues: 1-153 <LEH>
C;Comment: This myoglobin was isolated from skeletal muscle.
C;Superfamily: globin; globin homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F;2-147/Domain: globin homology <GLB>
F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 90.9%; Score 722; DB 1; Length 153;
Best Local Similarity 90.1%; Pred. No. 7.6e-53;
Matches 137; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDPRFKHLKTEAEKASEDL 61
Db 2 LSEAEWQLVHVWAKVEADLSGHGOEILIRLPKSHPTLEKFDPRFKHLKTEAEKASEDL 61
QY 62 KKHGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG 121
Db 62 KKHGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG 121
QY 122 DFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153
Db 122 DFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153

Search completed: August 10, 2004, 15:29:27
Job time : 7.66828 secs

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:19:00 ; Search time 4.81598 Seconds
(without alignments)
1654.229 Million cell updates/sec

Title: US-09-455-978b-76

Perfect score: 794

Sequence: 1 VLSEGEWQLVHLVWAKVEAD.....ALELFRKDIAAKYKELGYQG 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	794	100.0	153	1 MYG_PHYCA	P02185 physeter ca
2	773	97.4	153	1 MYG_KOGSI	P02184 kogia simus
3	746	94.0	153	1 MYG_ESCGI	P02177 eschrichtiu
4	740	93.2	153	1 MYG_MEGNO	P02178 megaptera n
5	738	92.9	153	1 MYG_INIGE	P02181 inia geoffr
6	737	92.8	153	1 MYG_GLOME	P02174 globicephal
7	734	92.4	153	1 MYG_ORCOR	P02173 orcinus orc
8	733	92.3	153	1 MYG_TURTR	P02172 tursiops tr
9	731	92.1	153	1 MYG_BALAC	P02179 ballaenopter
10	731	92.1	153	1 MYG_PROPH	P02176 phocoenode
11	726	91.4	153	1 MYG_BALPH	P02180 ballaenopter
12	725	91.3	153	1 MYG_MESCA	P02183 mesoplodon
13	722	90.9	153	1 MYG_CASFI	P14396 castor fibe
14	722	90.9	153	1 MYG_ZIPCA	P02182 ziphius cav
15	717	90.3	153	1 MYG_HORSE	P02188 equus cabal
16	707	89.0	153	1 MYG_PANTR	P02145 pan troglod
17	705	88.8	153	1 MYG_PREPO	P02166 perodicticu
18	703	88.5	153	1 MYG_RABIT	P02170 oryctolagus
19	701	88.3	153	1 MYG_GORBE	P02147 gorilla gor
20	701	88.3	153	1 MYG_ONDZI	P32428 ondatra zib
21	700	88.2	153	1 MYG_PIG	P02189 sus scrofa
22	699	88.0	153	1 MYG_HUMAN	P02144 homo sapien
23	695	87.5	153	1 MYG_MACFA	P02150 macaca fasc
24	694	87.4	153	1 MYG_PONPY	P02148 pongo pygma
25	693	87.3	153	1 MYG_HYLAG	P02146 tylobates a
26	693	87.3	153	1 MYG_TUPLG	P02165 tupiaia glis
27	691	87.0	153	1 MYG_AOTTR	P02151 aotus trivi
28	691	87.0	153	1 MYG_PAPAN	P02149 papio anubi
29	690	86.9	153	1 MYG_CALJA	P02152 callithrix
30	690	86.9	153	1 MYG_LEPMU	P02163 rousettus a
31	688	86.6	153	1 MYG_LEPMU	P02169 lepitemur m
32	688	86.6	153	1 MYG_LUTLU	P11343 lutra lutra
33	687	86.5	153	1 MYG_ORYAF	P02164 orycteropus

34	685	86.3	153	1 MYG_NYCCO	P02167 nycticebus
35	684	86.1	153	1 MYG_ZALCA	P02161 zalophus ca
36	683	86.0	153	1 MYG_CTEGU	P20856 ctenodactyl
37	682	85.9	153	1 MYG_HALGR	P02162 halichoerus
38	680	85.6	153	1 MYG_DELLE	P83682 delphinapte
39	679	85.5	153	1 MYG_GALCR	P02168 galago cras
40	678	85.4	153	1 MYG_CFBAP	P02153 cebus apell
41	678	85.4	153	1 MYG_LAGLA	P02154 lagothrix l
42	678	85.4	153	1 MYG_OCHPR	P02171 ochotona pr
43	677	85.3	153	1 MYG_SPAEH	P04248 spalax leuc
44	674	84.9	153	1 MYG_CEREL	P02191 cervus elap
45	674	84.9	153	1 MYG_PROGU	P04249 proechimys

ALIGNMENTS

RESULT 1
MYG_PHYCA
AC P02185; STANDARD; PRT; 153 AA.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myoglobin.
GN MB.
OS Physeter catodon (Sperm whale) (Physeter macrocephalus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;
OC Physetridae; Physeter.
OX NCBI_TaxID=9755;
RN [1]
RP SEQUENCE.
RC TISSUE=Heart muscle;
RA Edmundson A.B.;
RT "Amino-acid sequence of sperm whale myoglobin.";
RL Nature 205:883-887(1965).
RN [2]
RP REVISION TO 122.
RC TISSUE=Skeletal muscle;
RA Romero-Herrera A.E., Lehmann H.;
RT "Residue 122 of sperm whale and horse myoglobin.";
RL Biochim. Biophys. Acta 336:318-323(1974).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF METMYOGLOBIN.
RX MEDLINE=77144097; PubMed=845959;
RA Takano T.;
RT "Structure of myoglobin refined at 2.0-A resolution. I.
J. Mol. Biol. 110:537-568(1977).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF DEOXYMYOGLOBIN.
RX MEDLINE=77144098; PubMed=845960;
RA Takano T.;
RT "Structure of myoglobin refined at 2.0-A resolution. II. Structure of
deoxymyoglobin from sperm whale.";
J. Mol. Biol. 110:569-584(1977).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF OXYMYOGLOBIN.
RX MEDLINE=81119812; PubMed=7463482;
RA Phillips S.E.V.;
RT "Structure and refinement of oxymyoglobin at 1.6-A resolution.";
J. Mol. Biol. 142:531-554(1980).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=91132649; PubMed=1994031;
RA Lionetti C., Guanziroli M.G., Frigerio F., Ascenzi P., Bolognesi M.;
RT "X-ray crystal structure of the ferric sperm whale myoglobin:
imidazole complex at 2.0-A resolution.";
J. Mol. Biol. 217:409-412(1991).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX MEDLINE=99158603; PubMed=10049310;

RA Brunori M., Cutruzzola F., Savino C., Travaglini-Allocatelli C.,
RA Vallone B., Gibson Q.H.;
RT "Structural dynamics of ligand diffusion in the protein matrix: a
RT study on a new myoglobin mutant Y(B10) Q(E7) R(B10).";
RL Biophys. J. 76:1259-1269(1999).
RN [8].
RP X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS).
RX MEDLINE=20160904; PubMed=10681426;
RA Brunori M., Vallone B., Cutruzzola F., Travaglini-Allocatelli C.,
RA Berendzen J., Chu K., Sweet R.M., Schlichting I.;
RT "The role of cavities in protein dynamics: crystal structure of a
RT photocyclic intermediate of a mutant myoglobin";
RL Proc. Natl. Acad. Sci. U.S.A. 97:2058-2063(2000).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates
CC the movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A90591; MYWHP.
DR PDB; 101M; 08-APR-98.
DR PDB; 102M; 08-APR-98.
DR PDB; 103M; 08-APR-98.
DR PDB; 104M; 08-APR-98.
DR PDB; 105M; 08-APR-98.
DR PDB; 106M; 08-APR-98.
DR PDB; 107M; 08-APR-98.
DR PDB; 108M; 20-MAY-98.
DR PDB; 109M; 08-APR-98.
DR PDB; 110M; 08-APR-98.
DR PDB; 111M; 08-APR-98.
DR PDB; 112M; 08-APR-98.
DR PDB; 1A6G; 21-OCT-98.
DR PDB; 1A6K; 06-APR-99.
DR PDB; 1A6M; 06-APR-99.
DR PDB; 1A6N; 06-APR-99.
DR PDB; 1A6S; 01-APR-97.
DR PDB; 1AJG; 12-NOV-97.
DR PDB; 1AJH; 12-NOV-97.
DR PDB; 1BVC; 31-JUL-95.
DR PDB; 1BVD; 31-JUL-95.
DR PDB; 1BZ6; 11-NOV-98.
DR PDB; 1BZP; 10-MAY-99.
DR PDB; 1BZR; 10-MAY-99.
DR PDB; 1CH1; 09-APR-99.
DR PDB; 1CH2; 09-APR-99.
DR PDB; 1CH3; 09-APR-99.
DR PDB; 1CH5; 09-APR-99.
DR PDB; 1CH7; 09-APR-99.
DR PDB; 1CH9; 09-APR-99.
DR PDB; 1CIK; 09-APR-99.
DR PDB; 1CIO; 09-APR-99.
DR PDB; 1CO8; 14-JUN-99.
DR PDB; 1CO9; 14-JUN-99.
DR PDB; 1CP0; 14-JUN-99.
DR PDB; 1CP5; 14-JUN-99.
DR PDB; 1CPW; 14-JUN-99.
DR PDB; 1CO2; 16-AUG-99.
DR PDB; 1DO1; 02-APR-00.
DR PDB; 1DO3; 02-APR-00.
DR PDB; 1DO4; 02-APR-00.
DR PDB; 1DO7; 02-APR-00.
DR PDB; 1DTI; 30-JAN-00.
DR PDB; 1DTM; 14-JUN-00.
DR PDB; 1DUK; 08-APR-03.
DR PDB; 1DOO; 27-DEC-00.
DR PDB; 1DXC; 02-APR-00.
DR PDB; 1DXD; 20-JUL-00.
DR PDB; 1EBC; 13-AUG-99.
DR PDB; 1F63; 19-JUL-00.
DR PDB; 1F65; 19-JUL-00.
DR PDB; 1F6H; 21-JUL-00.
DR PDB; 1FCS; 31-OCT-93.
DR PDB; 1HJT; 12-NOV-97.
DR PDB; 1IOP; 08-APR-98.
DR PDB; 1IRC; 11-JUL-96.
DR PDB; 1JDO; 27-MAY-98.
DR PDB; 1JF6; 16-JAN-02.
DR PDB; 1JP8; 16-JAN-02.
DR PDB; 1JP9; 16-JAN-02.
DR PDB; 1JPB; 16-JAN-02.
DR PDB; 1JW8; 10-OCT-01.
DR PDB; 1L2K; 21-AUG-02.
DR PDB; 1LTW; 23-DEC-96.
DR PDB; 1MBC; 09-JAN-89.
DR PDB; 1MBD; 27-OCT-83.
DR PDB; 1MBI; 15-OCT-91.
DR PDB; 1MBN; 27-OCT-83.
DR PDB; 1MBO; 30-SEP-83.
DR PDB; 1MCY; 07-DEC-95.
DR PDB; 1MGN; 31-JAN-94.
DR PDB; 1MLF; 31-AUG-94.
DR PDB; 1MLG; 31-AUG-94.
DR PDB; 1MLH; 31-AUG-94.
DR PDB; 1MLJ; 31-AUG-94.
DR PDB; 1MLK; 31-AUG-94.
DR PDB; 1MLL; 31-AUG-94.
DR PDB; 1MLM; 31-AUG-94.
DR PDB; 1MLN; 31-AUG-94.
DR PDB; 1MLO; 31-AUG-94.
DR PDB; 1MLQ; 07-FEB-95.
DR PDB; 1MLR; 31-AUG-94.
DR PDB; 1MLS; 31-AUG-94.
DR PDB; 1MLU; 31-AUG-94.
DR PDB; 1MOA; 07-FEB-95.
DR PDB; 1MOB; 07-FEB-95.
DR PDB; 1MOC; 07-FEB-95.
DR PDB; 1MOD; 07-FEB-95.
DR PDB; 1MTI; 15-SEP-95.
DR PDB; 1MTJ; 15-SEP-95.
DR PDB; 1MTK; 15-SEP-95.
DR PDB; 1MYF; 27-FEB-95.
DR PDB; 1MYM; 31-JAN-94.
DR PDB; 1OBM; 08-APR-98.
DR PDB; 1OFJ; 27-MAY-98.
DR PDB; 1OFK; 11-NOV-98.
DR PDB; 1SPE; 08-MAR-96.
DR PDB; 1SWM; 31-JAN-94.
DR PDB; 1TES; 08-NOV-96.
DR PDB; 1VXA; 01-AUG-96.
DR PDB; 1VXB; 01-AUG-96.
DR PDB; 1VXC; 01-AUG-96.
DR PDB; 1VXD; 01-AUG-96.
DR PDB; 1VXE; 01-AUG-96.
DR PDB; 1VXF; 01-AUG-96.
DR PDB; 1VXG; 01-AUG-96.
DR PDB; 1VXH; 01-AUG-96.
DR PDB; 1YOH; 07-DEC-96.
DR PDB; 1YOI; 07-DEC-96.
DR PDB; 2CMM; 31-JAN-94.
DR PDB; 2MB5; 15-APR-91.
DR PDB; 2MBW; 23-DEC-96.
DR PDB; 2MGA; 31-JAN-94.
DR PDB; 2MGB; 31-JAN-94.
DR PDB; 2MGC; 31-JUL-94.
DR PDB; 2MGD; 31-JUL-94.
DR PDB; 2MGE; 31-JUL-94.
DR PDB; 2MGF; 31-JUL-94.
DR PDB; 2MGH; 31-JUL-94.
DR PDB; 2MGI; 31-JUL-94.
DR PDB; 2MGJ; 31-JUL-94.
DR PDB; 2MGK; 31-JUL-94.
DR PDB; 2MGL; 31-JUL-94.
DR PDB; 2MGW; 31-JUL-94.
DR PDB; 2MYA; 31-JAN-94.
DR PDB; 2MYB; 31-JAN-94.
DR PDB; 2MYC; 31-JAN-94.

```

Query Match          100.0%; Score 794; DB 1; Length 153;
Best Local Similarity 100.0%; Pred. No. 1.2e-61;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASED 60
DB 1 VLSGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASED 60
QY 61 LKKGVTVTALGAILKKKGHEAELEKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
DB 61 LKKGVTVTALGAILKKKGHEAELEKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
QY 121 GDFGADAQAGMANKALELFRKDIAAKYKELGYQG 153
DB 121 GDFGADAQAGMANKALELFRKDIAAKYKELGYQG 153

RESULT 2
MYG_KOGSI
AC P02184; STANDARD; PRT; 153 AA.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myoglobin.
GN MB.
OS Kogia simus (Dwarf sperm whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;
OC Physeteridae; Kogia.
OX NCBI_TaxID=9752;
RN [1]
RP SEQUENCE.
RX MEDLINE=77134684; PubMed=843520;
RA Dwulet F.E., Jones B.N., Lehman L.D., Gurd F.R.N.;
RT "The complete amino acid sequence of the major component myoglobin of
  dwarf sperm whale (Kogia simus).";
RL Biochemistry 16:873-877(1977).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
  movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A90407; MYWHW.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 93 93 IRON (HEME DISTAL LIGAND).
FT METAL 64 64 IRON (HEME PROXIMAL LIGAND).
FT METAL 93 93 IRON (HEME DISTAL LIGAND).
SQ SEQUENCE 153 AA; 17237 MW; 5771A432C7B32614 CRC64;

Query Match          97.4%; Score 773; DB 1; Length 153;
Best Local Similarity 96.7%; Pred. No. 7.5e-60;
Matches 148; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VLSGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASED 60
DB 1 VLSGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASED 60
QY 61 LKKGVTVTALGAILKKKGHEAELEKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
DB 61 LKKGVTVTALGAILKKKGHEAELEKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
QY 121 GDFGADAQAGMANKALELFRKDIAAKYKELGYQG 153
DB 121 ADFGADAQAGMANKALELFRKDIAAKYKELGYQG 153

RESULT 3
MYG_ESCGI
ID -MYG_ESCGI STANDARD; PRT; 153 AA.

```

```

AC P02177;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Eschrichtius gibbosus (California gray whale) (Eschrichtius robustus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Eschrichtiidae; Eschrichtius.
OX NCBI_TaxID=9764;
RN [1]
RP SEQUENCE.
RX TISSUE=Skeletal muscle;
RX MEDLINE=76232191; PubMed=938629;
RA Bogardt R.A. Jr., Dwulet F.E., Lehman L.D., Jones B.N., Gurd F.R.N.;
RT "Complete primary structure of the major component myoglobin of
  California gray whale (Eschrichtius gibbosus).";
RL Biochemistry 15:2597-2602(1976).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
  movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A02499; MYWHC.
DR HSSP; P02185; 1BZ6.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 93 93 IRON (HEME DISTAL LIGAND).
FT METAL 64 64 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17118 MW; 4A5587BA0E805C7D CRC64;

Query Match          94.0%; Score 746; DB 1; Length 153;
Best Local Similarity 92.8%; Pred. No. 1.6e-57;
Matches 142; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 VLSGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASED 60
DB 1 VLSDAEWQLVLIWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASED 60
QY 61 LKKGVTVTALGAILKKKGHEAELEKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
DB 61 LKKGVTVTALGAILKKKGHEAELEKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
QY 121 GDFGADAQAGMANKALELFRKDIAAKYKELGYQG 153
DB 121 GDFGADAQAGMANKALELFRKDIAAKYKELGYQG 153

RESULT 4
MYG_MEGNO
ID -MYG_MEGNO STANDARD; PRT; 153 AA.
AC P02178;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Megaptera novaeangliae (Humpback whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Megaptera.
OX NCBI_TaxID=9773;
RN [1]
RP SEQUENCE.
RX TISSUE=Skeletal muscle;
RX MEDLINE=79021601; PubMed=698193;
RA Lehman L.D., Dwulet F.E., Jones B.N., Bogardt R.A. Jr.,
  Krueckeberg S.T., Visscher R.B., Gurd F.R.N.;
RT "Complete amino acid sequence of the major component myoglobin from
  the humpback whale, Megaptera novaeangliae.";

```

```

RL Biochemistry 17:3736-3739 (1978).
RN [2]
RP SEQUENCE OF 1-60.
RX MEDLINE=68049168; PubMed=6059350;
RA Edman P., Begg G.;
RT "A protein sequenator.";
RL Eur. J. Biochem. 1:80-91(1967).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
CC movement of oxygen within muscles.
CC -!- MISCELLANEOUS: This sequence was the first determined using the
CC Edman sequencing system.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A90416; MYWHH.
DR HSSP; P02185; 1B26.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND)..
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17132 MW; 4FAFD81A0E805C7D CRC64;

Query Match 93.2%; Score 740; DB 1; Length 153;
Best Local Similarity 92.2%; Pred. No. 5.2e-57;
Matches 141; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKASD 60
DB 1 VLSDAEWQLVNIWAKVADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKASD 60

QY 61 LKKGVTVTALGAILKKKGHEAEKPLAQSHATKKKIPKYLEFISEAIHVLHSRHP 120
DB 61 LKKGNTVLTALGGILKKKGHEAEKPLAQSHATKKKIPKYLEFISDAIIVLHSRHP 120

QY 121 GDFGADAQGMNKALELFRKIDIAAKYKELGYQG 153
DB 121 ADFGADAQGMNKALELFRKIDIAAKYKELGFGQ 153

RESULT 5
ID MYG INICE STANDARD; PRT; 153 AA.
AC P02181;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myoglobin.
GN MB.
OS Inia geoffrensis (Amazon dolphin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Iniidae;
OC Inia.
OX NCBI_TaxID=9725;
RN [1]
RP SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=76062412; PubMed=1191640;
RA Dwulet F.E., Bogardt R.A., Jones B.N., Lehman L.D., Gurd F.R.N.;
RT "The complete amino acid sequence of the major component myoglobin of
RT Amazon river dolphin (Inia geoffrensis).";
RL Biochemistry 14:5336-5343(1975).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
CC movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR HSSP; P02188; 1WLA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.

```

```

KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND)..
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17071 MW; 72222ED24EBCE5D CRC64;

Query Match 92.9%; Score 738; DB 1; Length 153;
Best Local Similarity 91.4%; Pred. No. 7.7e-57;
Matches 139; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKASD 61
DB 2 LSDGEWQLVNIWKGKVEADLAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKASD 61

QY 62 KKHGVTVTALGAILKKKGHEAEKPLAQSHATKKKIPKYLEFISEAIHVLHSRHP 121
DB 62 KKHGNTVLTALGGILKKKGHEAEKPLAQSHATKKKIPKYLEFISEAIHVLHSRHP 121

QY 122 DFGADAQGMNKALELFRKIDIAAKYKELGYQG 153
DB 122 DFGADAQGMNKALELFRKIDIAAKYKELGFGH 153

RESULT 6
ID MYG GLOME STANDARD; PRT; 153 AA.
AC P02174;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Globicephala melaleuca (Long-finned pilot whale) (Globicephala melas).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;
OC Globicephala.
OX NCBI_TaxID=9731;
RN [1]
RP SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=78187213; PubMed=656376;
RA Jones B.N., Dwulet F.E., Lehman L.D., Garner M.H., Bogardt R.A. Jr.,
RA Garner W.H., Gurd F.R.N.;
RT "Complete amino acid sequence of myoglobin from the pilot whale,
RT Globicephala melaleuca.";
RL Biochemistry 17:1971-1974(1978).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
CC movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A02496; MYWHT.
DR HSSP; P02185; 1A6W.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND)..
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17085 MW; 510980DECB83DFD CRC64;

Query Match 92.8%; Score 737; DB 1; Length 153;
Best Local Similarity 92.1%; Pred. No. 9.4e-57;
Matches 140; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKASD 61
DB 2 LSDGEWQLVNIWKGKVEADLAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKASD 61

QY 62 KKHGVTVTALGAILKKKGHEAEKPLAQSHATKKKIPKYLEFISEAIHVLHSRHP 121
DB 62 KKHGNTVLTALGGILKKKGHEAEKPLAQSHATKKKIPKYLEFISEAIHVLHSRHP 121

QY 122 DFGADAQGMNKALELFRKIDIAAKYKELGYQG 153

```


QY 62 KKHGVTTLTALGAILKKKGHHEAEKPLAQSHATKHKPIKYLEFISEAIHVLHSRHPG 121
 Db 62 KKHGNTVLTALGAILKKKGHDAELKPLAQSHATKHKPIKYLEFISEAIHVLHSRHPA 121
 QY 122 DFGADAQAMNKALELFRKDIKAAKYKELGYQG 153
 Db 122 BFGADAQAMNKALELFRKDIKAAKYKELGFHG 153

RESULT 9

MYG_BALAC
 ID MYG_BALAC STANDARD; PRT; 153 AA.
 AC P02179;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myoglobin.
 GN MB.
 OS Balaenoptera acutorostrata (Minke whale) (Lesser rorqual).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
 OC Balaenopteridae; Balaenoptera.
 OX NCBI_TaxID=9767;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=77112438; PubMed=836810;
 RA Lehman L.D., Dwulet F.E., Bogardt R.A. Jr., Jones B.N., Gurd F.R.N.;
 RT "The complete amino acid sequence of the major component myoglobin
 from the arctic minke whale, Balaenoptera acutorostrata."
 RL Biochemistry 16:706-709(1977).
 CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
 movement of oxygen within muscles.
 CC -!- SIMILARITY: Belongs to the globin family.
 DR PIR; A02501; MYWK.
 DR HSSP; P02185; 1B26.
 DR InterPro; IPR000971; Globin.
 DR Pfam; PF00042; Myoglobin.
 DR PRINTS; PR00613; MYOGLOBIN.
 DR PROSITE; PS01033; GLOBIN; 1.
 KW Heme; Oxygen transport; Transport; Muscle.
 FT METAL 64 64 IRON (HEME DISTAL LIGAND).
 FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 153 AA; 17155 MW; A5364E71B9705C6E CRC64;

Query Match 92.1%; Score 731; DB 1; Length 153;
 Best Local Similarity 90.8%; Pred. No. 3.1e-56;
 Matches 139; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
 QY 1 VLSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKHLEKTEAEWKASED 60
 Db 1 VLSDAEHLVLIWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKHLEKTEAEWKASED 60
 QY 61 LKHGVTTLTALGAILKKKGHHEAEKPLAQSHATKHKPIKYLEFISEAIHVLHSRHP 120
 Db 61 LKHGNTVLTALGGILKKKGHHEAEKPLAQSHATKHKPIKYLEFISDAIHVLHSRHP 120
 QY 121 DFGADAQAMNKALELFRKDIKAAKYKELGYQG 153
 Db 121 AEFGADAQAMNKALELFRKDIKAAKYKELGFQG 153

RESULT 10

MYG_PHOPH
 ID MYG_PHOPH STANDARD; PRT; 153 AA.
 AC P02176;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myoglobin.
 GN MB.
 OS Phocoenoides phocoena (Harbor porpoise), and

OS Phocoenoides dalli dalli (Dall's porpoise).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Phocoenidae;
 OC Phocoena.
 OX NCBI_TaxID=9742, 9745;
 RN [1]
 RP SEQUENCE.
 RC SPECIES=P.phocoena;
 RX MEDLINE=69177451; PubMed=5782005;
 RA Bradshaw R.A., Gurd F.R.N.;
 RT "Comparison of myoglobins from harbor seal, porpoise, and sperm
 whale. V. The complete amino acid sequences of harbor seal and
 porpoise myoglobins."
 RL J. Biol. Chem. 244:2167-2181(1969).
 RN [2]
 RP SEQUENCE, AND REVISIONS TO 83 AND 85.
 RC SPECIES=P.phocoena, and P.D.dalli;
 RX MEDLINE=79000346; PubMed=687594;
 RA Meuth J.L., Jones B.N., Garner W.H., Gurd F.R.N.;
 RT "Complete amino acid sequence of the myoglobin from the Dall porpoise
 (Phocoenoides dalli dalli) and reinvestigation of the primary
 structure of the myoglobin from common porpoise (Phocoena
 phocoena)."
 RL Biochemistry 17:3429-3431(1978).
 CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
 movement of oxygen within muscles.
 CC -!- SIMILARITY: Belongs to the globin family.
 DR PIR; B90415; MYPD.
 DR PIR; B92045; MYPE.
 DR HSSP; P02185; 1B26.
 DR InterPro; IPR000971; Globin.
 DR Pfam; PF00042; Myoglobin.
 DR PRINTS; PR00613; MYOGLOBIN.
 DR PROSITE; PS01033; GLOBIN; 1.
 KW Heme; Oxygen transport; Transport; Muscle.
 FT METAL 64 64 IRON (HEME DISTAL LIGAND).
 FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 153 AA; 17101 MW; 3FA2F4561A35CC2E CRC64;
 Query Match 92.1%; Score 731; DB 1; Length 153;
 Best Local Similarity 90.8%; Pred. No. 3.1e-56;
 Matches 138; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
 QY 2 LSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKHLEKTEAEWKASEDL 61
 Db 2 LSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKHLEKTEAEWKASEDL 61
 QY 62 KKHGVTTLTALGAILKKKGHHEAEKPLAQSHATKHKPIKYLEFISEAIHVLHSRHPG 121
 Db 62 KKHGNTVLTALGGILKKKGHDAELKPLAQSHATKHKPIKYLEFISEAIHVLHSRHPA 121
 QY 122 DFGADAQAMNKALELFRKDIKAAKYKELGYQG 153
 Db 122 BFGADAQAMNKALELFRKDIATKYKELGFHG 153
 RESULT 11
 MYG_BALPH
 ID MYG_BALPH STANDARD; PRT; 153 AA.
 AC P02180;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myoglobin.
 GN MB.
 OS Balaenoptera physalus (Finback whale) (Common rorqual).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
 OC Balaenopteridae; Balaenoptera.
 OX NCBI_TaxID=9770;
 RN [1]
 RP SEQUENCE.

```

RC TISSUE=Skeletal muscle;
RX MEDLINE=78187212; PubMed=656375;
RA Dimarchi R.D., Wang C.-C., Hemenway J.B., Gurd F.R.N.;
RT "Complete amino acid sequence of the major component myoglobin of
RL finback whale (Balaenoptera physalus).";
RL Biochemistry 17:1968-1970(1978).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
CC movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
CC PIR; A02502; MYWHF.
DR HSSP; P02185; 1BZ6.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND).
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17086 MW; 8AFC89C1BF00291F CRC64;

Query Match 91.4%; Score 726; DB 1; Length 153;
Best Local Similarity 90.2%; Pred. No. 8.4e-56;
Matches 138; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 VLSGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFRFKHLKTEAEKASDDL 60
Db 1 VLTDAEHLVNLINWAKVEADVAGHGQDILISLFKGPETLEKFRFKHLKTEAEKASDDL 60

Qy 61 LKHGVTVTALGAILKKKGHHEAEKPLAQSHATKHKIKPIKYLEFISEAIIHVLHSRHP 120
Db 61 LKHGNTVLTALGGILKKKGHHEAEKPLAQSHATKHKIKPIKYLEFISDAIIHVLHSRHP 120

Qy 121 GDFGADAQAGAMNKALELFRKDIKAAKYKELGYQG 153
Db 121 ADFGADAQAGAMNKALELFRKDIKAAKYKELGFQG 153

RESULT 12
MYG_MESCA
ID MYG_MESCA STANDARD; PRT; 153 AA.
AC P02183;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Mesoplodon carlhubbsi (Hubb's whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Ziphiidae;
OC Mesoplodon.
OX NCBI_taxID=9758;
RN [1]
RP SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=81000592; PubMed=7407230;
RA Dwalet J.A., Dwalet F.E., Gurd F.R.N.;
RT "Complete amino acid sequence of the major component myoglobin from
RL Hubbs's beaked whale, Mesoplodon carlhubbsi.";
RL Biochim. Biophys. Acta 624:121-129(1980).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
CC movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
CC PIR; A02505; MYWHU.
DR HSSP; P02185; 1A6M.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND).
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).

```

```

SQ SEQUENCE 153 AA; 17136 MW; 8BA3DB535CBE91E9 CRC64;

Query Match 91.3%; Score 725; DB 1; Length 153;
Best Local Similarity 90.1%; Pred. No. 1e-55;
Matches 137; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Qy 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFRFKHLKTEAEKASDDL 61
Db 2 LSEABWQLVHVWAKVEADLSGHQGEILIRLFKGPETLEKFRFKHLKTEAEKASDDL 61

Qy 62 KKHGVTVTALGAILKKKGHHEAEKPLAQSHATKHKIKPIKYLEFISEAIIHVLHSRHP 121
Db 62 KKHGHTVLTALGGILKKKGHHEAEKPLAQSHATKHKIKPIKYLEFISDAIIHVLHSKHP 121

Qy 122 DFGADAQAGAMNKALELFRKDIKAAKYKELGYQG 153
Db 122 DFGADAQAGAMTKALELFRKDIKAAKYKELGFHG 153

RESULT 13
MYG_CASFI
ID MYG_CASFI STANDARD; PRT; 153 AA.
AC P14396;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Castor fiber (Eurasian beaver).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Castoridae; Castor.
OX NCBI_taxID=10185;
RN [1]
RP SEQUENCE.
RA Sukhomlinov B.F., Drobot L.B.;
RT "The primary structure of myoglobin from beaver (Castor fiber); II
RT peptic peptides of the tryptic insoluble core. Reconstruction of
RT beaver myoglobin polypeptide chain.";
RL Bioorg. Khim. 6:542-546(1980).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
CC movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; JN0410; JN0410.
DR HSSP; P02189; 1MWC.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND).
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17020 MW; 4FD93C4E116B6D4D CRC64;

Query Match 90.9%; Score 722; DB 1; Length 153;
Best Local Similarity 87.5%; Pred. No. 1.8e-55;
Matches 133; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

Qy 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFRFKHLKTEAEKASDDL 61
Db 2 LSDGEWQLVHVWGKVEADLAGHQEVILIRLFKGPETLEKFNKFKHKSDEMKASDDL 61

Qy 62 KKHGVTVTALGAILKKKGHHEAEKPLAQSHATKHKIKPIKYLEFISEAIIHVLHSRHP 121
Db 62 KKHGVTVLTALGGVLKKKGHHEAEIKPLAQSHATKHKIKPIKYLEFISEAIIHVLQSKHP 121

Qy 122 DFGADAQAGAMNKALELFRKDIKAAKYKELGYQG 153
Db 122 BFGADAGAGAMNKALELFRKDIKAAKYKELGFQG 153

RESULT 14
MYG_ZIPCA

```

```

ID MYG_ZIPCA STANDARD; PRT; 153 AA.
AC P02182;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Ziphys cavirostris (Goose-beaked whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Ziphiidae;
OC Ziphys.
OX NCBI_TaxID=9760;
RN [1]_
RP SEQUENCE.
RX MEDLINE=81063290; PubMed=7437458;
RA Lehman L.D., Jones B.N., Dwulet F.E., Bogardt R.A. Jr., Gurd F.R.N.;
RT "Complete amino acid sequence of the major component myoglobin from
the goose-beaked whale, Ziphys cavirostris.";
RL Biochim. Biophys. Acta 625:221-229(1980).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A02504; MYWZ.
DR HSP; P02185; IAGM.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND).
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17178 MW; 8DC3BD451EDAA1E9 CRC64;

Query Match 90.9%; Score 722; DB 1; Length 153;
Best Local Similarity 90.1%; Pred. No. 1.8e-55;
Matches 137; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 2 LSEGEWQVLVHWAKVEADVAGHGQDILIRLPKSHPTLEKEDRPFKHLKTEAEKASDGL 61
Db 2 LSEAEWQVLVHWAKVEADLSHGQGEILIRLPKSHPTLEKEDRPFKHLKTEAEKASDGL 61
QY 62 KKHGTVTLTALGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISEAILHVLHSRHPG 121
Db 62 KKHGTVTLTALGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISEAILHVLHSRHPG 121
QY 122 DFGADAQAMNKALEFRKDIAAKYKELGYQG 153
Db 122 DFGADAQAMNKALEFRKDIAAKYKELGFGHG 153

RESULT 15
ID MYG_HORSE STANDARD; PRT; 153 AA.
AC P02183;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myoglobin.
GN MB.
OS Equus caballus (Horse), and
OS Equus burchelli (Plains zebra) (Equus quagga).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796, 9790;
RN [1]_
RP SEQUENCE.
RC SPECIES=Horse; TISSUE=Heart muscle;
RX MEDLINE=70064092; PubMed=4902609;
RA Dautrevaux M., Boulanger Y., Han K., Biserte G.;
RT "Covalent structure of horse myoglobin.";
RL Eur. J. Biochem. 111:267-277(1969).
RN [2]_

```

```

RP SEQUENCE.
RC SPECIES=Horse; TISSUE=Skeletal muscle;
RA Romero-Herrera A.E., Lehmann H.;
RT "Residue 122 of sperm whale and horse myoglobin.";
RL Biochim. Biophys. Acta 336:318-323(1974).
RN [3]_
RP SEQUENCE OF 1-15 AND 56-70.
RC SPECIES=Horse;
RX MEDLINE=90147691; PubMed=2302197;
RA Jahn W., Ward L.D., Reid G.E., Moritz R.L., Simpson R.J.;
RT "Internal amino acid sequencing of proteins by in situ cyanogen
bromide cleavage in polyacrylamide gels.";
RL Biochem. Biophys. Res. Commun. 166:139-145(1990).
RN [4]_
RP COMPOSITION OF TRYPTIC AND PEPTIC PEPTIDES.
RC SPECIES=E.burchelli;
RX MEDLINE=75184110; PubMed=1095063;
RA Darbre P.D., Romero-Herrera A.E., Lehmann H.;
RT "Comparison of the myoglobin of the zebra (Equus burchelli) with that
of the horse (Equus caballus).";
RL Biochim. Biophys. Acta 393:201-204(1975).
RN [5]_
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RC SPECIES=Horse;
RX MEDLINE=90294302; PubMed=2359126;
RA Evans S.V., Brayer G.D.;
RT "High-resolution study of the three-dimensional structure of horse
heart metmyoglobin.";
RL J. Mol. Biol. 213:885-897(1990).
RN [6]_
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RC SPECIES=Horse;
RX MEDLINE=88153746; PubMed=3346247;
RA Evans S.V., Brayer G.D.;
RT "Horse heart metmyoglobin. A 2.8-A resolution three-dimensional
structure determination.";
RL J. Biol. Chem. 263:4263-4268(1988).
RN [7]_
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF MUTANT THR-64.
RC SPECIES=Horse;
RX MEDLINE=95383313; PubMed=7654702;
RA Bogumil R., Maurus R., Hildebrand D.P., Brayer G.D., Mauk A.G.;
RT "Origin of the pH-dependent spectroscopic properties of
pentacoordinate metmyoglobin variants.";
RL Biochemistry 34:10483-10490(1995).
RN [8]_
RP X-RAY CRYSTALLOGRAPHY (1.45 ANGSTROMS).
RC SPECIES=Horse;
RX MEDLINE=20168811; PubMed=10706294;
RA Chu K., Voitchovsky J., McMahon B.H., Sweet R.M., Berendzen J.,
Schlichting I.;
RT "Structure of a ligand-binding intermediate in wild-type carbonmonoxy
myoglobin.";
RL Nature 403:921-923(2000).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A90603; MYH0Z.
DR PIR; A91098; MYH0.
DR PDB; 1YMA; 31-JAN-94.
DR PDB; 1YMB; 31-JAN-94.
DR PDB; 1YMC; 31-JAN-94.
DR PDB; 1HRM; 26-JAN-95.
DR PDB; 1HSY; 27-FEB-95.
DR PDB; 1RSE; 23-DEC-96.
DR PDB; 1XCH; 17-SEP-97.
DR PDB; 1WLA; 14-JAN-98.
DR PDB; 1BJE; 28-JAN-98.
DR PDB; 1AZI; 25-FEB-98.
DR PDB; 1DWR; 03-MAR-00.
DR PDB; 1DWS; 24-JUN-03.
DR PDB; 1DWT; 03-MAR-00.
DR PDB; 1GJN; 01-MAR-02.

```

```
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle; 3D-structure.
FT METAL 64 64
FT METAL 93 93
FT METAL 122 122
FT CONFLICT 122 122
FT HELIX 4 18
FT TURN 19 20
FT HELIX 21 35
FT HELIX 37 40
FT TURN 41 42
FT TURN 44 48
FT HELIX 52 57
FT HELIX 59 77
FT TURN 78 80
FT HELIX 83 94
FT TURN 95 96
FT HELIX 101 118
FT TURN 120 121
FT HELIX 125 148
FT TURN 149 150
SQ SEQUENCE 153 AA; 16951 MW; 89CA01974231E93C CRC64;

Query Match          90.3%; Score 717; DB 1; Length 153;
Best Local Similarity 88.2%; Pred. No. 5e-55;
Matches 134; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKASBDL 61
Db 2 LSDGEWQVLNVWGVKEADVAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKASBDL 61

QY 62 KKHGTVTLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPG 121
Db 62 KKHGTVTLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPG 121

QY 122 DFGADAQGMNKALELFRKDIAAKYKELGYQG 153
Db 122 DFGADAQGMNKALELFRNDIAAKYKELGFQG 153
```

Search completed: August 10, 2004, 15:26:50
Job time : 5.81598 secs

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:22:05 ; Search time 18.3378 Seconds
(without alignments)
2632.505 Million cell updates/sec

Title: US-09-455-978b-76
Perfect score: 794
Sequence: 1 VLSEGEWQLVHLHWAKVEAD.....ALELFRKDIAAKYKELGYQG 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	647	81.5	154	11	Q9QZ76
2	460	57.9	99	4	Q8WHV6
3	321.5	40.5	147	13	Q9DGI1
4	312.5	39.4	147	13	Q9DGI0
5	305.5	38.5	147	13	Q9DGI7
6	288.5	36.3	147	13	Q9DGI2
7	288	36.3	146	13	Q9DGI8
8	285.5	36.0	142	13	Q9DWH9
9	283.5	35.7	147	13	Q7T044
10	271.5	34.2	141	13	P79846
11	267.5	33.7	147	13	Q9DGI9
12	237.5	29.9	110	13	Q98963
13	215.5	27.1	190	11	Q8BN80
14	211.5	26.6	190	4	Q8N2X5
15	194.5	24.5	103	13	Q8AY74
16	192.5	24.2	103	13	O57410

17	169.5	21.3	177	13	Q802S6
18	146	18.4	147	6	Q9BEI1
19	144.5	18.2	143	13	Q7ZT21
20	143.5	18.1	143	13	Q7SZV8
21	143	18.0	142	13	Q9PVL9
22	143	18.0	147	6	Q9BEI2
23	142.5	17.9	143	13	Q8UV18
24	142	17.9	147	11	Q88752
25	140	17.6	144	13	Q8UW52
26	138.5	17.4	147	11	Q9D0B2
27	138	17.4	146	11	Q9QUN8
28	138	17.4	147	11	Q9CR49
29	137	17.3	35	6	Q865L4
30	136.5	17.2	143	13	Q8UW95
31	135.5	17.1	143	13	Q8JIN0
32	135.5	17.1	143	13	Q8AYQ1
33	134.5	16.9	143	13	Q7SZV9
34	134	16.9	142	6	Q95233
35	134	16.9	142	11	Q9CWS5
36	133.5	16.8	143	13	Q803Z5
37	133	16.8	142	11	Q91V15
38	133	16.8	147	6	Q9GLX5
39	132	16.6	142	6	Q28743
40	132	16.6	142	13	Q90ZAS
41	132	16.6	147	6	Q9GJS7
42	132	16.6	147	6	Q9GLX7
43	132	16.6	147	13	Q90938
44	131.5	16.6	143	13	Q8JH83
45	131.5	16.6	143	13	Q8JH82

ALIGNMENTS

RESULT 1

Q9QZ76 PRELIMINARY; PRT; 154 AA.

AC Q9QZ76;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Myoglobin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_taxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RA Santos R.A., Giannocco G., Poyares L.L., Nunes M.T.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF197916; AAF05848.1; -
DR HSP; P02189; IMWC.
DR GO; GO:0005344; F:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 154 AA; 17157 MW; 98B31BE57FA020DA CRC64;

Query Match 81.5%; Score 647; DB 11; Length 154;

Best Local Similarity 78.9%; Pred. No. 2.1e-52;

Matches 120; Conservative 18; Mismatches 14; Indels 0; Gaps 0;

Qy 2 LSEGEWQLVHLHWAKVADVAGHGDILIRLFKSHPETLEKDFPKHLKTEAEMKASDL 61

Db 3 LSDGEWQVNLINGKVEGLAGHGOEVLISLFKAHPETLEKDFKHLKSEEMKSSDL 62

Qy 62 KKHGVTVLTALGAILKKKGHEALKPLAQSHATKHKIPKYLEFISEAIIHLVLSRHPG 121

```
Db 63 KHGCTVLTALGTLKKGQRAEIQPLAQSHATKHKIPVKYLEFISEVILQVLKRYSG 122
Qy 122 DFGADAQAMNKALELFRKDIKAAKYKELGYQG 153
Db 123 DFGADAQAMNKALELFRNDIAAKYKELGFGQ 154

RESULT 2
Q8WH6 PRELIMINARY; PRT; 99 AA.
AC Q8WH6;
DT 01-WAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Similar to myoglobin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RA Strauberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC018001; RAH18001.1; -.
DR HSP; P02185; 111M.
DR GO; GO:0005344; F:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 99 AA; 10870 MW; 73E67F581F79C479 CRC64;

Query Match 57.9%; Score 460; DB 4; Length 99;
Best Local Similarity 87.9%; Pred. No. 3e-35;
Matches 87; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy 55 MKASDLKKHGVTVLTALGAILKKGGHAEELKPLAQSHATKHKIPVKYLEFISEBAILHV 114
Db 1 MKASDLKKHGVTVLTALGAILKKGGHAEELKPLAQSHATKHKIPVKYLEFISECTIOV 60

Qy 115 LHSRHPGDFGADAQAMNKALELFRKDIKAAKYKELGYQG 153
Db 61 LQSKHPGDFGADAQAMNKALELFRKDWASNYKELGFGQ 99

RESULT 3
Q9DGJ1 PRELIMINARY; PRT; 147 AA.
AC Q9DGJ1;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-WAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Myoglobin.
OS Makaira nigricans (Blue marlin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Scombroidei;
OC Xiphidae; Makaira.
OX NCBI_TaxID=13604;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
RT "Oxygen affinity and amino acid sequence of myoglobins from endothermic and ectothermic fish.";
RL Am. J. Physiol. Regul. Integr. Comp. Physiol. 280:R1123-R1133(2001).

Query Match 39.4%; Score 312.5; DB 13; Length 147;
Best Local Similarity 44.6%; Pred. No. 2.5e-21;
Matches 66; Conservative 21; Mismatches 58; Indels 3; Gaps 2;

Qy 6 EQWLVLHVWAKVEADVAGHGDILIRLPKSHPETLEKDFRPHKLTAEAMKASEDLKKHG 65
Db 3 DFDVILKFWGPVEADYTHGNLVLTRELPETQKLPKFAGI-AKADMAGNAAISAHG 61
```

```
DR EMBL; AF291833; AAG02107.1; -.
DR HSP; P02205; 1MYT.
DR GO; GO:0005344; F:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 147 AA; 15842 MW; F52D010973F4D84B CRC64;

Query Match 40.5%; Score 321.5; DB 13; Length 147;
Best Local Similarity 46.7%; Pred. No. 3.7e-22;
Matches 70; Conservative 20; Mismatches 53; Indels 7; Gaps 3;

Qy 6 EQWLVLHVWAKVEADVAGHGDILIRLPKSHPETLEKDFRPHKLTAEAMKASEDLKKHG 65
Db 3 DFEVLKHWGPVEADYTHGNLVLTRELPETQKLPKFAGI-AKADMAGNAAISAHG 61

Qy 66 VTVLTAALGAILKKGGHAEELKPLAQSHATKHKIPVKYLEFISEBAILHVLSRHPGDFGA 125
Db 62 ATVLKKGELKAKGSHAAIKPMANSHATKHKIPKPFELISEVIGVMHEK---AGL 117

Qy 126 DAQG--AMNKALELFRKDIKAAKYKELGYQG 153
Db 118 DAQGAKALKNVWTTIIADIEANYKELGFTG 147

RESULT 4
Q9DGJ0 PRELIMINARY; PRT; 147 AA.
AC Q9DGJ0;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Myoglobin.
OS Sarda chiliensis (Sard).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Scombroidei;
OC Scombridae; Sarda.
OX NCBI_TaxID=8231;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
RT "Oxygen affinity and amino acid sequence of myoglobins from endothermic and ectothermic fish.";
RL Am. J. Physiol. Regul. Integr. Comp. Physiol. 280:R1123-R1133(2001).
DR EMBL; AF291834; AAG02108.1; -.
DR HSP; P02205; 1MYT.
DR GO; GO:0005344; F:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 147 AA; 15797 MW; D0864510E730506 CRC64;

Query Match 39.4%; Score 312.5; DB 13; Length 147;
Best Local Similarity 44.6%; Pred. No. 2.5e-21;
Matches 66; Conservative 21; Mismatches 58; Indels 3; Gaps 2;

Qy 6 EQWLVLHVWAKVEADVAGHGDILIRLPKSHPETLEKDFRPHKLTAEAMKASEDLKKHG 65
Db 3 DFDVILKFWGPVEADYTHGNLVLTRELPETQKLPKFAGI-AQADMAGNAAISAHG 61
```



```

QY 66 VTTLTALGAILKKKGHEAEALPKAQSHATKHKIPKYLEFISEAIIHVLHSRHPDGA 125
Db 62 ATVLKGLGELLKAGNHAAILKPMANSHATKHKIPINNFKLISEIIVKVMQEKAGDAG- 120

QY 126 DAQGMNKALELFRKIDIAKYKELGYQG 153
Db 121 -GQALRNVMGIIADLEANYKELGFSG 147

RESULT 5
Q9DGI7 PRELIMINARY; PRT; 147 AA.
AC Q9DGI7;
RC TISSUE=Skeletal muscle;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Myoglobin.
OS Thunnus albacares (Yellowfin tuna) (Neothunnus macropterus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC Scombridae; Thunnus.
OX NCBI_TaxID=8236;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
RT "Oxygen affinity and amino acid sequence of myoglobins from
endothermic and ectothermic fish.";
RL Am. J. Physiol. Regul. Integr. Comp. Physiol. 280:R1123-R1133(2001).
DR EMBL; AF291838; AAG02112.1; -.
DR PIR; A02526; MYTUY.
DR HSSP; P02205; 1MYT.
DR GO; GO:0005344; F:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR PRINTS; PR00613; Myoglobin.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 147 AA; 15660 MW; 3D13814403B48D44 CRC64;

Query Match 38.5%; Score 305.5; DB 13; Length 147;
Best Local Similarity 43.9%; Pred. No. 1.1e-20;
Matches 65; Conservative 22; Mismatches 58; Indels 3; Gaps 2;

QY 6 EWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDLKKHG 65
Db 3 DFDAVLKCGPVEADYTTMGVLVTLRPFKEHPTQKLPFKFAGI-AQADIAGNAAISAHG 61

QY 66 VTTLTALGAILKKKGHEAEALPKAQSHATKHKIPKYLEFISEAIIHVLHSRHPDGA 125
Db 62 ATVLKGLGELLKAGSHAAILKPMANSHATKHKIPINNFKLISEIIVKVMQEKAGDAG- 120

QY 126 DAQGMNKALELFRKIDIAKYKELGYQG 153
Db 121 -GQALRNVMGIIADLEANYKELGFSG 147

RESULT 6
Q9DGJ2 PRELIMINARY; PRT; 147 AA.
AC Q9DGJ2;
RC TISSUE=Skeletal muscle;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Myoglobin.
OS Thunnus alalunga (Albacore).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

```

```

OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC Scombridae; Thunnus.
OX NCBI_TaxID=8235;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
RT "Oxygen affinity and amino acid sequence of myoglobins from
endothermic and ectothermic fish.";
RL Am. J. Physiol. Regul. Integr. Comp. Physiol. 280:R1123-R1133(2001).
DR EMBL; AF291832; AAG02106.1; -.
DR HSSP; P02205; 1MYT.
DR GO; GO:0005344; F:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR PRINTS; PR00613; Myoglobin.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 147 AA; 15667 MW; 26114A4B946A1191 CRC64;

Query Match 36.3%; Score 288.5; DB 13; Length 147;
Best Local Similarity 41.2%; Pred. No. 4.4e-19;
Matches 61; Conservative 25; Mismatches 59; Indels 3; Gaps 2;

QY 6 EWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDLKKHG 65
Db 3 DFDAVLKCGPVEADYTTMGVLVTLRPFKEHPTQKLPFKFAGI-AQADLAGNAAISAHG 61

QY 66 VTTLTALGAILKKKGHEAEALPKAQSHATKHKIPKYLEFISEAIIHVLHSRHPDGA 125
Db 62 ATVLKGLGELLKAGSHAAILKPMANSHATKHKIPINNFKLISEIIVKVMQEKAGDAG- 120

QY 126 DAQGMNKALELFRKIDIAKYKELGYQG 153
Db 121 -GQALRNVMGIIADLEANYKELGFTG 147

RESULT 7
Q9DGI8 PRELIMINARY; PRT; 146 AA.
AC Q9DGI8;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Myoglobin.
OS Katsuwonus pelamis (Skipjack tuna) (Bonito).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC Scombridae; Katsuwonus.
OX NCBI_TaxID=8226;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
RT "Oxygen affinity and amino acid sequence of myoglobins from
endothermic and ectothermic fish.";
RL Am. J. Physiol. Regul. Integr. Comp. Physiol. 280:R1123-R1133(2001).
DR EMBL; AF291837; AAG02111.1; -.
DR HSSP; P02205; 1MYT.
DR GO; GO:0005344; F:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR PRINTS; PR00613; Myoglobin.
DR PROSITE; PS01033; GLOBIN; 1.

```

```
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 146 AA; 15396 MW; 307DAA2C6D9PDC27 CRC64;

Query Match 36.3%; Score 288; DB 13; Length 146;
Best Local Similarity 43.8%; Pred. No. 4.8e-19;
Matches 64; Conservative 21; Mismatches 53; Indels 8; Gaps 3;

QY 10 VLHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASEDLKKHGTVVL 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 VLKCGAVEADFNVTGGVILARLFKDHPTQKLPFKFAGI--TGDIAAGNAAVAAGHATVVL 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 TALGAILKKKGHHEALKPLAQSHATKHKIPIKYLEFISEAIIHVLHSHRHPGDFGADAQG 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 KKLSELLKAKGNHAAIIPLANSHAKQKIPINNFKLITEALHVLHEK----AGLDAAG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 130 --AMNKALELFRKDIKAAKYLEGYOG 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 QTALRNVMGIIVADLEANYKELGFTG 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
Q90WH9 PRELIMINARY; PRT; 142 AA.
AC Q90WH9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Myoglobin (Fragment).
OS Hemipterus americanus (Sea raven).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Scorpaeniformes;
OC Cottoidae; Hemipteridae; Hemitripterus.
OX NCBI_TaxID=8094;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart ventricle;
RA Grove T.J., Sidell B.D.;
RT "Myoglobin from sea raven, Hemitripterus americanus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY029587; AAK49781.1; -.
DR HSSP; P02185; 111M.
DR GO; GO:0005344; F:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
FT NON_TER 1
SQ SEQUENCE 142 AA; 15111 MW; 2DB1EF4602F929D3 CRC64;

Query Match 36.0%; Score 285.5; DB 13; Length 142;
Best Local Similarity 44.2%; Pred. No. 7.9e-19;
Matches 65; Conservative 18; Mismatches 57; Indels 7; Gaps 3;

QY 9 LVLHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASEDLKKHGTVV 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 LVLKCGPVEADYAAVGSVLTRLFTEHPDTQKLPFKFAGI-AQGDMAADAGISAHGATV 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 69 LTAIGAILKKKGHHEALKPLAQSHATKHKIPIKYLEFISEAIIHVLHSHRHPGDFCADAQ 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 LRLGELLKAKGSHAAILKPLANSHATKHKIPINNFRLLITEVIGKVGMEK----TGLDAA 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 129 G--AMNKALELFRKDIKAAKYLEGYOG 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 GQALRNVMVAIVADMEADYKLGFTG 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
Q7T044 PRELIMINARY; PRT; 147 AA.
AC Q7T044;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Myoglobin.
OS Channichthys rhinoceratus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Notothenioidae; Channichthyidae; Channichthys.
OX NCBI_TaxID=70438;
RN [1]
RP SEQUENCE FROM N.A.
RA Grove T.J., Hendrickson J.W., Sidell B.D.;
RT "Myoglobin Gene Sequence from the Antarctic Notothenioid, Channichthys rhinoceratus.";
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY341058; AAK17544.1; -.
DR EMBL; AY341058; AAK17544.1; -.
SQ SEQUENCE 147 AA; 15695 MW; 256D05AA3E2DCF76 CRC64;

Query Match 35.7%; Score 283.5; DB 13; Length 147;
Best Local Similarity 42.6%; Pred. No. 1.3e-18;
Matches 63; Conservative 22; Mismatches 56; Indels 7; Gaps 3;

QY 6 EMQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASEDLKKHG 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 DFDVLMKCGPVEADHATHGSLVTRLFTEHPDTQKLPFKFAGI-ANGDLADGAGVSAHG 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 66 VTVALGAILKKKGHHEALKPLAQSHATKHKIPIKYLEFISEAIIHVLHSHRHPGDFGA 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 ATVLKKGDLKARGGHAALLPLSSSHATKHKIPINFKLIAEVIGKVGMEK----AGL 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 126 DAQG--AMNKALELFRKDIKAAKYLEGY 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 DAAGQTALRNVMVAIITDMEADYKELGF 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
P79846 PRELIMINARY; PRT; 141 AA.
AC P79846;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Myoglobin (Fragment).
OS Pagetopsis macropterus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Notothenioidae; Channichthyidae; Pagetopsis.
OX NCBI_TaxID=36194;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart ventricle;
RA Vayda M.E., Yuan M.-L., Small D.J., Costello L., Sidell B.D.;
RT "Extreme conservation of the myoglobin sequence among Antarctic notothenioid fishes.";
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U71152; AAB41673.1; -.
DR HSSP; P02205; 1MYT.
DR GO; GO:0005344; F:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
FT NON_TER 1
SQ SEQUENCE 141 AA; 14869 MW; 8EDF8F2F23210E80 CRC64;
```

```

Query Match          34.2%; Score 271.5; DB 13; Length 141;
Best Local Similarity 42.3%; Pred. No. 1.6e-17;
Matches 60; Conservative 20; Mismatches 59; Indels 3; Gaps 2;

QY 10 VLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFRPKHLKTEAEWKASEDLKKHGVTYL 69
Db 1 VLKWCMPMEADYATHGGLVTLRFTHEHPETLKLFPKFGI-AHGDLAGDAGVSAHGATVL 59
QY 70 TALGAILKKKGHEAEKPLAQSHATKHPIKYLEFISEAIIHVLHSHRPGDFGADAQG 129
Db 60 NKGDLKARGAHAALLKPLSSSHATKHPIPIINFKLIAEVIGKVMEEKAGVDAG-GQT 117

QY 130 ANKALELFRKIOIAKYKELGY 151
Db 118 GLRNVMAVIIADMEAYKELGF 139

RESULT 11
ID Q9DGI9 PRELIMINARY; PRT; 147 AA.
AC Q9DGI9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Myoglobin.
OS Scomber japonicus (Chub mackerel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Scombroidei;
OC Scombridae; Scomber.
OX NCBI_TaxID=13676;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
RT "Oxygen affinity and amino acid sequence of myoglobins from
endothermic and ectothermic fish.";
RL Am. J. Physiol. Regul. Integr. Comp. Physiol. 280:R1123-R1133(2001).
DR EMBL; AF291835; AAG02109.1; -.
DR HSSP; P02205; IMYT.
DR GO; GO:0005344; P:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 147 AA; 15766 MW; 44EB9A4611EE0366 CRC64;

Query Match          33.7%; Score 267.5; DB 13; Length 147;
Best Local Similarity 39.9%; Pred. No. 3.9e-17;
Matches 59; Conservative 23; Mismatches 63; Indels 3; Gaps 2;

QY 6 EWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFRPKHLKTEAEWKASEDLKKHG 65
Db 3 DFDVAVLKFVGPVEADYDKIGNVLTRLTFTHEPDQKLPKFGI-GL-GDMAGNAISAHG 61
QY 66 VTVLTALGAILKKKGHEAEKPLAQSHATKHPIKYLEFISEAIIHVLHSHRPGDFGDA 125
Db 62 ATVLKKALEVLKAGNAGIIRPLANSATKHKIINNFKLITELIVKVMQEKAGLDAG- 120

QY 126 DAQGANNAKALELFRKIDIAKYKELGYQG 153
Db 121 -GQTALRNVMGVFIADMDANYKELGFSG 147

RESULT 12
Q98963
ID Q98963 PRELIMINARY; PRT; 110 AA.
AC Q98963;

```

```

DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Myoglobin (fragment).
OS Notothenia coriiceps (black rockcod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes;
OC Nototheniidae; Nototheniidae; Notothenia.
OX NCBI_TaxID=8208;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart ventricle;
RA Vayda M.E., Small D.J., Sidell B.D.;
RT "Expression of the myoglobin gene in Antarctic channichthyid fishes.";
RL (in) Battaglia B., Valencia J., Walton S.W.H. (eds.);
VI SCAR ANTARCTIC COMMUNITIES, pp.1-1, Cambridge University Press,
Cambridge, UK (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart ventricle;
RX MEDLINE=98360008; PubMed=9694664;
RA Small D.J., Vayda M.E., Sidell B.D.;
RT "A novel vertebrate myoglobin gene containing three A+T-rich introns
is conserved among Antarctic teleost species which differ in myoglobin
expression.";
RL J. Mol. Evol. 47:156-166(1998).
DR EMBL; U68350; AAC69245.1; -.
DR HSSP; P02205; IMYT.
DR GO; GO:0005344; P:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR Pfam; PF00042; Globin; 1.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
FT NON_TER 1
FT NON_TER 110
SQ SEQUENCE 110 AA; 11630 MW; B43B5865F0E6708F CRC64;

Query Match          29.9%; Score 237.5; DB 13; Length 110;
Best Local Similarity 45.9%; Pred. No. 1.7e-14;
Matches 50; Conservative 15; Mismatches 43; Indels 1; Gaps 1;

QY 10 VLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFRPKHLKTEAEWKASEDLKKHGVTYL 69
Db 1 VLKWCMPMEADYATHGGLVTLRFTHEHPETLKLFPKFGI-AHGDLAGDAGVSAHGATVL 59
QY 70 TALGAILKKKGHEAEKPLAQSHATKHPIKYLEFISEAIIHVLHSHR 118
Db 60 NKGDLKARGAHAALLKPLSSSHATKHPIPIINFKLIAEVIGKVMEEK 108

RESULT 13
Q98N80
ID Q98N80 PRELIMINARY; PRT; 190 AA.
AC Q98N80;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Mus musculus cytoglobin.
CN CYGB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eye;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of

```

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:18:30 ; Search time 31.8547 Seconds
(without alignments)
1632.058 Million cell updates/sec

Title: US-09-455-978b-77
Perfect score: 933
Sequence: 1 MSNDNDTLVTADVRNGIDGH.....DELVARFLPMLKLTFFDQOI 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	933	100.0	489	4	AAE04641	Aae04641 Halobacte
2	203	21.8	39	4	AAE04677	Aae04677 Haem prot
3	142	15.2	432	4	AAE04642	Aae04642 Bacillus
4	99	10.6	448	1	AAE070493	Aap70493 Protein G
5	99	10.6	448	1	AAE095030	Aap95030 Protein G
6	99	10.6	448	2	AAE07013	Aar07013 Protein G
7	99	10.6	448	2	AAE10004	Aar10004 Streptoco
8	99	10.6	448	2	AAE53290	Aar53290 Streptoco
9	99	10.6	1175	2	AAE34570	Aay34570 Porphorym
10	99	10.6	1232	2	AAE34569	Aay34569 Porphorym
11	99	10.6	1266	2	AAE34568	Aay34568 Porphorym
12	99	10.6	1269	2	AAE34428	Aay34428 Porphorym
13	98	10.5	593	2	AAE07014	Aar07014 Protein G
14	98	10.5	593	2	AAE62944	Aar62944 Streptoco
15	98	10.5	594	2	AAE10005	Aar10005 Streptoco
16	89	9.5	955	2	AAE57365	Aar57365 K39 polyp
17	89	9.5	955	2	AAE03691	Aaw03691 Leishmani
18	88	9.4	505	6	ABU38101	Abu38101 Protein e
19	87	9.3	496	6	ABU37440	Abu37440 Protein e
20	87	9.3	505	6	ABP79309	Abp79309 N. gonorr
21	87	9.3	523	3	AAE34454	Aay34454 Porphorym
22	87	9.3	533	2	AAE34330	Aay34330 Porphorym
23	86.5	9.3	318	3	AAE07677	Aab07677 Amino aci
24	86	9.2	483	7	ADC31377	Adc31377 Human rov
25	85.5	9.2	302	6	ADA33559	Ada33559 Acinetoba

ALIGNMENTS

RESULT 1

AAE04641	ID	AAE04641	standard; protein; 489 AA.
XX	AC	AAE04641;	
XX	AC	AAE04641;	
DT	11-SEP-2003	(revised)	
DT	04-SEP-2001	(first entry)	
XX	DE	Halobacterium salinarum HemAT-Hs protein.	
XX	KW	Haem binding protein; HemAT-Hs; HemAT-Bs; gaseous ligand sensor;	
KW	KW	oxygen storage; artificial photosynthesis; signalling function;	
KW	KW	alpha-haemoglobin; myoglobin; therapy.	
XX	OS	Halobacterium salinarum.	
XX	PN	WO200140475-A2.	
XX	PD	07-JUN-2001.	
XX	PF	05-DEC-2000; 2000WO-US033048.	
XX	PR	06-DEC-1999; 99US-00455978.	
XX	PA	(UYHA-) UNIV HAWAII.	
XX	PI	Alam M, Larsen R;	
XX	DR	WPI; 2001-374832/39.	
XX	DR	N-PSDB; AAD08991.	
XX	PT	Novel isolated bacterial heme binding protein, HemAT-Hs or HemAT-Bs which	
PT	PT	reversibly binds oxygen with low affinity, useful for controlled storage	
PT	PT	of oxygen and for sensing gaseous ligands such as oxygen.	
XX	PS	Claim 6; Page 10; 94pp; English.	
XX	CC	The present invention relates to isolated archeal and bacterial haem	
CC	CC	binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen with	
CC	CC	low affinity. Haem binding protein is useful for controlled storage of	
CC	CC	oxygen by allowing haem binding protein to bind and store oxygen, and	
CC	CC	triggering the release of oxygen from haem binding protein by activating	
CC	CC	the signalling domain. Haem binding protein is useful for sensing gaseous	
CC	CC	ligands such as oxygen, NO, CO, or CN. Blood substitute comprising haem	
CC	CC	binding protein is useful for treating a patient suffering from low blood	
CC	CC	levels by administering and regulating the oxygen binding of the haem-	
CC	CC	binding protein by modifying the signalling domain. Haem binding protein	

CC is useful for haem-based catalysis, for artificial photosynthesis and for
 CC identifying potential signalling functions of mutated alpha-haemoglobin
 CC and myoglobin causing several diseases. The present sequence is
 CC Halobacterium salinarum HemAT-Hs protein which is salt tolerant.
 CC (Updated on 11-SEP-2003 to standardise OS field)

XX Sequence 489 AA;

Query Match 100.0%; Score 933; DB 4; Length 489;
 Best Local Similarity 100.0%; Pred. No. 2.6e-89;
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNDNTLTADVRNGIDGHALADRGIDLEDAEIAWLSFTGIDDDTMAALAEQPLFEAT 60
 DB 1 MSNDNTLTADVRNGIDGHALADRGIDLEDAEIAWLSFTGIDDDTMAALAEQPLFEAT 60
 QY 61 ADALVTDFDHLSEYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAARARIG 120
 DB 61 ADALVTDFDHLSEYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAARARIG 120
 QY 121 KHDVLGLGPDVYLGAITYRYTGLDADLADDDVADRGEEAAAADDELVARFLPMLKLITF 180
 DB 121 KHDVLGLGPDVYLGAITYRYTGLDADLADDDVADRGEEAAAADDELVARFLPMLKLITF 180

QY 181 DQOI 184

DB 181 DQOI 184

RESULT 2

AAE04677

ID AAE04677 standard; peptide; 39 AA.

XX

AC AAE04677;

XX

DT 11-SEP-2003 (revised)

DT 04-SEP-2001 (first entry)

XX

DE Haem protein related Halobacterium salinarum protein fragment.

XX

KW Haem binding protein; HemAT-Hs; HemAT-Bs; gaseous ligand sensor;

KW oxygen storage; artificial photosynthesis; signalling function;

KW alpha-haemoglobin; myoglobin; therapy.

XX

OS Halobacterium salinarum.

XX

FH Key Location/Qualifiers

FT Domain 19..39

FT /label= M2_Box

XX

FN WO200140475-A2.

XX

PD 07-JUN-2001.

XX

PF 05-DEC-2000; 2000WO-US033048.

XX

PR 06-DEC-1999; 99US-00455978.

XX

PA (UYHA-) UNIV HAWAII.

XX

PI Alam M, Larsen R;

XX

DR WPI; 2001-374832/39.

XX

XX Novel isolated bacterial heme binding protein, HemAT-Hs or HemAT-Bs which

PT reversibly binds oxygen with low affinity, useful for controlled storage

PT of oxygen and for sensing gaseous ligands such as oxygen.

XX

PS Disclosure; Page 16; 94pp; English.

XX

XX The present invention relates to isolated archeal and bacterial haem

CC binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen with

CC low affinity. Haem binding protein is useful for controlled storage of

CC oxygen by allowing haem binding protein to bind and store oxygen, and
 CC triggering the release of oxygen from haem binding protein by activating
 CC the signalling domain. Haem binding protein is useful for sensing gaseous
 CC ligands such as oxygen, NO, CO, or CN. Blood substitute comprising haem
 CC binding protein is useful for treating a patient suffering from low blood
 CC levels by administering and regulating the oxygen binding of the haem-
 CC binding protein by modifying the signalling domain. Haem binding protein
 CC is useful for haem-based catalysis, for artificial photosynthesis and for
 CC identifying potential signalling functions of mutated alpha-haemoglobin
 CC and myoglobin causing several diseases. The present sequence is a haem
 CC protein related Halobacterium salinarum protein fragment. (Updated on 11
 CC -SEP-2003 to standardise OS field)

XX Sequence 39 AA;

Query Match 21.8%; Score 203; DB 4; Length 39;

Best Local Similarity 100.0%; Pred. No. 6e-14;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 QAEYLLGLGRGEYDTEYAARARIGIKIHDLVGLGPDVYL 134

DB 1 QAEYLLGLGRGEYDTEYAARARIGIKIHDLVGLGPDVYL 39

RESULT 3

AAE04642

ID AAE04642 standard; protein; 432 AA.

XX

AC AAE04642;

XX

DT 04-SEP-2001 (first entry)

XX

DE Bacillus subtilis HemAT-Bs protein.

XX

KW Haem binding protein; HemAT-Hs; HemAT-Bs; gaseous ligand sensor;

KW oxygen storage; artificial photosynthesis; signalling function;

KW alpha-haemoglobin; myoglobin; therapy.

XX

OS Bacillus subtilis.

XX

PN WO200140475-A2.

XX

PD 07-JUN-2001.

XX

PF 05-DEC-2000; 2000WO-US033048.

XX

PR 06-DEC-1999; 99US-00455978.

XX

PA (UYHA-) UNIV HAWAII.

XX

PI Alam M, Larsen R;

XX

DR WPI; 2001-374832/39.

XX

DR N-PSDB; AAD08992.

XX

XX Novel isolated bacterial heme binding protein, HemAT-Hs or HemAT-Bs which

PT reversibly binds oxygen with low affinity, useful for controlled storage

PT of oxygen and for sensing gaseous ligands such as oxygen.

XX

PS Claim 8; Page 11; 94pp; English.

XX

XX The present invention relates to isolated archeal and bacterial haem

CC binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen with

CC low affinity. Haem binding protein is useful for controlled storage of

CC oxygen by allowing haem binding protein to bind and store oxygen, and

CC triggering the release of oxygen from haem binding protein by activating

CC the signalling domain. Haem binding protein is useful for sensing gaseous

CC ligands such as oxygen, NO, CO, or CN. Blood substitute comprising haem

CC binding protein is useful for treating a patient suffering from low blood

CC levels by administering and regulating the oxygen binding of the haem-

CC binding protein by modifying the signalling domain. Haem binding protein

CC is useful for haem-based catalysis, for artificial photosynthesis and for

CC identifying potential signalling functions of mutated alpha-haemoglobin

CC and myoglobin causing several diseases. The present sequence is Bacillus
 CC subtilis HemaT-Bs protein
 CC
 XX
 SQ Sequence 432 AA;

Query Match 15.2%; Score 142; DB 4; Length 432;
 Best Local Similarity 25.3%; Pred. No. 4.8e-06;
 Matches 39; Conservative 29; Mismatches 76; Indels 10; Gaps 2;

QY 31 AEIAMLSTGIDDDTMAALAAEQPLFEATADALVDFYDHLSEYRTQDLFANSTKTVE 90
 Db 32 ADVKKQLKVRGLGDAELYVLEQLPLQENIYIVDAFYKNDHSSLMIDII-NHSSVD 90
 QY 91 QLKETAQEVLLGLRGREYDTEVAARIGKTHDVLGDPVYLGAYTRYTYTGLLDALAD 150
 Db 91 RLKQTLKRHIQEMFAGVIDDFIEKRNRIASIHRLGILLPKWYMGAFQELLSSMIDIV-- 148
 QY 151 DVVADRGEAAAAVDELVARPLPKLTLTFDQOI 184
 Db 149 -----EASITNQELLKAIKATTKILNLEOQL 175

RESULT 4
 AAP70493
 ID AAP70493 standard; protein; 448 AA.
 XX
 AC AAP70493;
 XX
 XX 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 06-MAR-1991 (first entry)
 DE Protein G.
 XX
 XX Protein G; antibody; Fc receptor.
 KW Streptococcus sp; 'Lancefield Group G strain'.
 OS
 XX
 FH Key Location/Qualifiers
 FT Active-site 228..297
 FT /label= active site B1
 FT Active-site 298..352
 FT /label= active site B2
 FT
 XX WO8705025-A.
 PN
 XX
 XX 27-AUG-1987.
 PD
 XX
 PF 17-FEB-1987; 87WO-US000329.
 PF
 XX
 PR 14-FEB-1986; 86US-00829354.
 PR 23-APR-1986; 86US-00854887.
 PR
 XX (GENX) GENEX CORP.
 XX (FAHN/) FAHNSTOCK S R.
 PA
 PA Fahnstock SR;
 PI
 XX
 XX WPI; 1987-250197/35.
 DR N-PSDB; AAN70811.
 DR
 XX Cloned Protein G gene - used for producing Protein G for detection and
 PT purificn. of antibodies and treatment of diseases.
 PT
 XX Disclosure; Fig 3; 68pp; English.
 XX
 XX Protein G expressed by inserting the gene into an expression vector. A
 CC second vector may also be used as a cryptic helper plasmid to stably
 CC maintain the first plasmid in the host cell. Bacterial Fc receptors such
 CC as Protein G can be used to detect and purify antibodies, and in the
 CC treatment of disease. Fc receptors are useful to purify antibodies to be
 CC used in the purificn. of protein drugs and as therapeutics. High levels
 CC of Protein G can be obt'd. in conditions favourable for isolation, using a

CC non-pathogenic host. Suitable cloning vectors are lambda gtl1, M13mp9 and
 CC pCX1066. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR
 CC -2003 to correct PA field.) (Updated on 25-MAR-2003 to correct PI field.)
 CC (Updated on 27-AUG-2003 to correct OS field.)
 XX
 SQ Sequence 448 AA;

Query Match 10.6%; Score 99; DB 1; Length 448;
 Best Local Similarity 25.3%; Pred. No. 0.17;
 Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;

QY 10 TADVRNGIDGHADRIGLDEAEIAWRLSFTGIDD-----DTMAALAAE----- 53
 Db 41 TPIIRNGGE---LTNLGNSETTLALRNEESATDLTAAAVADTVAAAAAENAGAAAAEAA 97
 QY 54 -----QPLFEATADAL-----VTDFYDHLSEYRTQDLFANSTKTVEQLKETAQEVLLG 102
 Db 98 AAAADALAKAKADALKFKNYGVSDYKNNL-----INNAKTVEGIDKLQAQV-- 145
 QY 103 LGRGEYDTEYAAQRARIKIHDLVGLGDPVYLGAYTRYTYTGLLDALADDDVADRGEAAA 162
 Db 146 -----ESAKKARISEATD-----GLSDFLKSQTPA---EDTVK 175

QY 163 AVDELVARPL 172
 Db 176 SIELAEAKVL 185

RESULT 5
 AAP95030
 ID AAP95030 standard; protein; 448 AA.
 XX
 AC AAP95030;
 XX
 XX 25-MAR-2003 (revised)
 DT 04-JUL-1990 (first entry)
 DE Protein G.
 XX
 XX Protein G; immunoglobulin; Fc receptor; ds.
 KW Streptococcus sp.
 OS
 XX WO8810306-A.
 PN
 XX
 XX 29-DEC-1988.
 PD
 XX 20-JUN-1988; 88WO-US002084.
 PF
 XX
 PR 19-JUN-1987; 87US-00063959.
 PR (GENX) GENEX CORP.
 PA (PHAA) PHARMACIA AB.
 PA
 XX Fahnstock SR;
 PI
 XX
 XX WPI; 1989-023848/03.
 DR N-PSDB; AAN91093.
 DR
 XX Cloned protein G variant genes - expressing proteins having
 PT immunoglobulin-binding properties of protein G and derived from
 PT Streptococcus sp.
 PT
 XX Disclosure; Page; 116pp; English.
 XX
 XX Protein G of non-pathogenic streptococcus and variants may be isolated,
 CC useful as bacterial Fc receptors eg in purification and detection of Abs.
 CC screening of hybridoma clones and treatment of disease. (Updated on 25-
 CC MAR-2003 to correct PA field.)
 XX
 SQ Sequence 448 AA;

Query Match 10.6%; Score 99; DB 1; Length 448;
 Best Local Similarity 25.3%; Pred. No. 0.17;
 Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;


```
XX SQ Sequence 448 AA;
Query Match 10.6%; Score 99; DB 2; Length 448;
Best Local Similarity 25.3%; Pred. No. 0.17;
Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;
QY 10 TADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDD-----DTMAALAAE----- 53
Db 41 TPIIRNGGE--LTNLLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAAWEEA 97
QY 54 -----QPLFEATADAL-----VTDPYDHLESYERTQDLFANSTKIVEOLKETQAEYLLG 102
Db 98 AAAADALAKAKADAKALKEFNKYGVSDYKKNL-----INNAKTVEGIKDLQAVV-- 145
QY 103 LGRGEYDTEYAAQRARIKIHVDVLGDPDVLGAVTRYTYTGLLDALADDDVADRGEAA 162
Db 146 -----ESAKKARISEATD-----GLSDFLKSTQTPA---EDTVK 175
QY 163 AVDELVARFL 172
Db 176 SIELAEAKVL 185

RESULT 8
AAARS3290
ID AARS3290 standard; protein; 448 AA.
XX AC AARS3290;
XX DT 16-OCT-2003 (revised)
XX DT 25-MAR-2003 (revised)
XX DT 06-JAN-1995 (first entry)
XX DE Streptococcus Protein G derived from strain GX7809.
XX KW Streptococcus Protein G; variant; IgG binding activity; immunoglobulin;
XX KW Lancefield Group G; bacterial Fc receptor.
XX OS Streptococcus sp. GX7805.
XX FH Key Location/Qualifiers
FT Active-site 228..282
FT Region 283..297
FT Active-site /note= "linking region"
FT 298..352
FT /label= B2
XX US5312901-A.
XX PD 17-MAY-1994.
XX PF 21-APR-1992; 92US-00871539.
XX PR 14-FEB-1986; 86US-00829354.
XX PR 23-APR-1986; 86US-00854887.
XX PR 19-JUN-1987; 87US-00063959.
XX PR 20-JUN-1988; 88US-00209236.
XX PR 19-JUN-1990; 90US-00540169.
XX PA (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX PI Farnestock SR;
XX DR WPI; 1994-159179/19.
XX DR N-PSDB; AAQ64644.
XX PT New recombinant streptococcal protein G variants - useful for antibody
XX PT detection and purification and for therapy.
XX PS Example 2; Fig 3 and Fig 8; 49pp; English.
```

```
XX CC A 1.9kb HindIII fragment containing the entire coding sequence for
CC Protein G was isolated from Streptococcus GX7809. The Protein G has IgG-
CC binding activity which has been localised to the B repeating structure.
CC Streptococcal protein G variants comprising the B domains are claimed.
CC (Updated on 25-MAR-2003 to correct PF field.) (Updated on 16-OCT-2003 to
CC standardise OS field)
XX SQ Sequence 448 AA;
Query Match 10.6%; Score 99; DB 2; Length 448;
Best Local Similarity 25.3%; Pred. No. 0.17;
Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;
QY 10 TADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDD-----DTMAALAAE----- 53
Db 41 TPIIRNGGE--LTNLLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAAWEEA 97
QY 54 -----QPLFEATADAL-----VTDPYDHLESYERTQDLFANSTKIVEOLKETQAEYLLG 102
Db 98 AAAADALAKAKADAKALKEFNKYGVSDYKKNL-----INNAKTVEGIKDLQAVV-- 145
QY 103 LGRGEYDTEYAAQRARIKIHVDVLGDPDVLGAVTRYTYTGLLDALADDDVADRGEAA 162
Db 146 -----ESAKKARISEATD-----GLSDFLKSTQTPA---EDTVK 175
QY 163 AVDELVARFL 172
Db 176 SIELAEAKVL 185

RESULT 9
AAAY34570
ID AAY34570 standard; protein; 1175 AA.
XX AC AAY34570;
XX DT 27-AUG-2003 (revised)
XX DT 20-MAR-2003 (revised)
XX DT 25-AUG-1999 (first entry)
XX DE Porphyromonas gingivalis protein PG9.
XX KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis; vaccine;
XX KW antigenic.
XX OS Porphyromonas gingivalis.
XX PN WO9929870-A1.
XX PD 17-JUN-1999.
XX PF 10-DEC-1998; 98WO-AU001023.
XX PR 10-DEC-1997; 97AU-00000839.
XX PR 31-DEC-1997; 97AU-00001182.
XX PR 30-JAN-1998; 98AU-00001546.
XX PR 10-MAR-1998; 98AU-00002264.
XX PR 09-APR-1998; 98AU-00002911.
XX PR 23-APR-1998; 98AU-00003128.
XX PR 05-MAY-1998; 98AU-00003338.
XX PR 22-MAY-1998; 98AU-00003654.
XX PR 29-JUL-1998; 98AU-00004917.
XX PR 30-JUL-1998; 98AU-00004963.
XX PR 04-AUG-1998; 98AU-00005028.
XX PA (CSLUC-) CSL LTD.
XX PI Ross BC, Barr IG, Patterson MA, Agius CT, Rothel LJ, Margetts MB;
XX PI Hocking DM, Webb EA;
XX DR WPI; 1999-385613/32.
XX DR N-PSDB; AAX91788.
```

XX Antigenic Porphyromonas gingivalis peptides for preventing gingivitis.
 PT Claim 1; Page 565-567; 588pp; English.
 XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAX34318 to
 CC AAX34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis. (Updated on 20-MAR-2003 to correct PR field.)
 CC (Updated on 27-AUG-2003 to correct OS field.)
 XX Sequence 1175 AA;
 SQ Query Match 10.6%; Score 99; DB 2; Length 1175;
 Best Local Similarity 25.6%; Pred. No. 0.66;
 Matches 46; Conservative 23; Mismatches 47; Indels 64; Gaps 9;
 QY 16 GIDGHALADRI-----GLDEAEIATWRLSFTG-----IDDDTMAALAAEQPL---- 56
 DB 333 GLDGLASLTRLRLRNQISKLEGLDRKLVRLKLDVSGNDIQSIDDIKLLAPILEQTLEKL 392
 QY 57 -----FEATADALVTDFYDHLSEYRTQDLFANSTKTVEQLKETQAEY-----LLG- 102
 DB 393 RIHDNPFVASSGILISPDVNH-----PEIKALLEKEKEKOKTSVEYHPFCVMLLGN 446
 QY 103 ---LGR---GEYDTEYAAQARIGIKHVDVGL-----GPDVILGAYTRYT 142
 DB 447 HSSGKTFLSQYDNTNYTQ-----KNTHVLSIHRSNPNNAIFYDFGGQDYHGIYQAFPT 501
 RESULT 10
 AAY34569
 ID AAY34569 standard; protein; 1232 AA.
 XX AC AAY34569;
 XX 27-AUG-2003 (revised)
 DT 20-MAR-2003 (revised)
 DT 25-AUG-1999 (first entry)
 XX DE Porphyromonas gingivalis protein PG9.
 XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis; vaccine;
 KW antigenic.
 XX Porphyromonas gingivalis.
 OS WO9929870-A1.
 XX PN 17-JUN-1999.
 XX PD 10-DEC-1998; 98WO-AU001023.
 XX PF 10-DEC-1997; 97AU-00000839.
 PR 31-DEC-1998; 97AU-00001182.
 PR 30-JAN-1998; 98AU-00001546.
 PR 10-MAR-1998; 98AU-00002264.
 PR 09-APR-1998; 98AU-00002911.
 PR 23-APR-1998; 98AU-00003128.
 PR 05-MAY-1998; 98AU-00003338.
 PR 22-MAY-1998; 98AU-00003654.
 PR 29-JUL-1998; 98AU-00004917.
 PR 30-JUL-1998; 98AU-00004963.
 PR 04-AUG-1998; 98AU-00005028.
 XX (CSLC-) CSL LTD.
 XX PA Ross BC, Barr IG, Patterson MA, Agius CT, Rothel LJ, Margetts MB;
 XX PI

PI Hocking DM, Webb EA;
 XX WPI: 1999-385613/32.
 DR N-PSDB; AAX91787.
 XX
 PT Antigenic Porphyromonas gingivalis peptides for preventing gingivitis.
 XX Claim 1; Page 563-565; 588pp; English.
 XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAX34318 to
 CC AAX34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis. (Updated on 20-MAR-2003 to correct PR field.)
 CC (Updated on 27-AUG-2003 to correct OS field.)
 XX Sequence 1232 AA;
 SQ Query Match 10.6%; Score 99; DB 2; Length 1232;
 Best Local Similarity 25.6%; Pred. No. 0.71;
 Matches 46; Conservative 23; Mismatches 47; Indels 64; Gaps 9;
 QY 16 GIDGHALADRI-----GLDEAEIATWRLSFTG-----IDDDTMAALAAEQPL---- 56
 DB 390 GLDGLASLTRLRLRNQISKLEGLDRKLVRLKLDVSGNDIQSIDDIKLLAPILEQTLEKL 449
 QY 57 -----FEATADALVTDFYDHLSEYRTQDLFANSTKTVEQLKETQAEY-----LLG- 102
 DB 450 RIHDNPFVASSGILISPDVNH-----PEIKALLEKEKEKOKTSVEYHPFCVMLLGN 503
 QY 103 ---LGR---GEYDTEYAAQARIGIKHVDVGL-----GPDVILGAYTRYT 142
 DB 504 HSSGKTFLSQYDNTNYTQ-----KNTHVLSIHRSNPNNAIFYDFGGQDYHGIYQAFPT 558
 RESULT 11
 AAY34568
 ID AAY34568 standard; protein; 1266 AA.
 XX AC AAY34568;
 XX 27-AUG-2003 (revised)
 DT 20-MAR-2003 (revised)
 DT 25-AUG-1999 (first entry)
 XX DE Porphyromonas gingivalis protein PG9.
 XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis; vaccine;
 KW antigenic.
 XX Porphyromonas gingivalis.
 OS WO9929870-A1.
 XX PN 17-JUN-1999.
 XX PD 10-DEC-1998; 98WO-AU001023.
 XX PF 10-DEC-1997; 97AU-00000839.
 PR 31-DEC-1998; 97AU-00001182.
 PR 30-JAN-1998; 98AU-00001546.
 PR 10-MAR-1998; 98AU-00002264.
 PR 09-APR-1998; 98AU-00002911.
 PR 23-APR-1998; 98AU-00003128.
 PR 05-MAY-1998; 98AU-00003338.
 PR 22-MAY-1998; 98AU-00003654.
 PR 29-JUL-1998; 98AU-00004917.
 PR 30-JUL-1998; 98AU-00004963.
 PR 04-AUG-1998; 98AU-00005028.

```

XX PA (CSLC-) CSL LTD.
XX PR Ross BC, Barr IG, Patterson MA, Agius CT, Rothel LJ, Margetts MB;
XX PI Hocking DM, Webb EA;
XX PT WPI; 1999-385613/32.
XX DR N-PSDB; AAX91786.
XX PT Antigenic Porphyromonas gingivalis peptides for preventing gingivitis.
XX PS Claim 1; Page 560-563; 588pp; English.
XX CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
XX CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX CC activity with a vaccine mechanism of action. The PG polypeptides can be
XX CC used as vaccines especially against Porphyromonas gingivalis. Probes can
XX CC be used to detect Porphyromonas gingivalis in standard hybridisation
XX CC assays. Porphyromonas gingivalis is involved in periodontal disease
XX CC especially gingivitis. (Updated on 20-MAR-2003 to correct PR field.)
XX CC (Updated on 27-AUG-2003 to correct OS field.)
XX SQ Sequence 1266 AA;

Query Match 10.6%; Score 99; DB 2; Length 1266;
Best Local Similarity 25.6%; Pred. No. 0.74;
Matches 46; Conservative 23; Mismatches 47; Indels 64; Gaps 9;

QY 16 GIDGHALADRI-----GLDEAEIAWRLSFTG-----IDDDTMAALAAEQPL--- 56
Db 424 GLDGLASLTSLRRNQISKLEGLRLKVLKLVSGNDIQSIDIKLLAPILEQTLEKL 483
QY 57 -----FEATADALVTDYFDHLESYERTQDLFANSTKTVQKETOAEY-----LLG- 102
Db 484 RIHDPFPVASSGLILSPYDNHL-----PEIKALLEKEKEKQKTSVEYHPFCKVMLLGN 537
QY 103 --LGR---GEYDTEYAAQARIGIKHIVLGI-----GPDVILGATRYTT 142
Db 538 HSSGKTTFLSQYDNTYTYQ-----KNTHVLSIHRSNPNNAIFYDFGGQDYHYGIYQAFET 592

RESULT 12
AAY34428
ID AAY34428 standard; protein; 1269 AA.
XX AC AAY34428;
XX DT 27-AUG-2003 (revised)
XX DT 20-MAR-2003 (revised)
XX DT 25-AUG-1999 (first entry)
XX DE Porphyromonas gingivalis protein PG9.
XX KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis; vaccine;
XX KW antigenic.
XX OS Porphyromonas gingivalis.
XX PN WO9929870-A1.
XX PD 17-JUN-1999.
XX PF 10-DEC-1998; 98WO-AU001023.
XX PR 10-DEC-1997; 97AU-00000839.
XX PR 31-DEC-1997; 97AU-00001182.
XX PR 30-JAN-1998; 98AU-00001546.
XX PR 10-MAR-1998; 98AU-00002264.
XX PR 09-APR-1998; 98AU-00002911.
XX PR 23-APR-1998; 98AU-00003128.
XX PR 05-MAY-1998; 98AU-00003338.

```

```

PR 22-MAY-1998; 98AU-00003654.
PR 29-JUL-1998; 98AU-00004917.
PR 30-JUL-1998; 98AU-00004963.
PR 04-AUG-1998; 98AU-00005028.
XX PA (CSLC-) CSL LTD.
XX PR Ross BC, Barr IG, Patterson MA, Agius CT, Rothel LJ, Margetts MB;
XX PI Hocking DM, Webb EA;
XX XX WPI; 1999-385613/32.
XX DR N-PSDB; AAX91646.
XX XX Antigenic Porphyromonas gingivalis peptides for preventing gingivitis.
XX PS Claim 1; Page 401-403; 588pp; English.
XX CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
XX CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX CC activity with a vaccine mechanism of action. The PG polypeptides can be
XX CC used as vaccines especially against Porphyromonas gingivalis. Probes can
XX CC be used to detect Porphyromonas gingivalis in standard hybridisation
XX CC assays. Porphyromonas gingivalis is involved in periodontal disease
XX CC especially gingivitis. (Updated on 20-MAR-2003 to correct PR field.)
XX CC (Updated on 27-AUG-2003 to correct OS field.)
XX SQ Sequence 1269 AA;

Query Match 10.6%; Score 99; DB 2; Length 1269;
Best Local Similarity 25.6%; Pred. No. 0.74;
Matches 46; Conservative 23; Mismatches 47; Indels 64; Gaps 9;

QY 16 GIDGHALADRI-----GLDEAEIAWRLSFTG-----IDDDTMAALAAEQPL--- 56
Db 427 GLDGLASLTSLRRNQISKLEGLRLKVLKLVSGNDIQSIDIKLLAPILEQTLEKL 486
QY 57 -----FEATADALVTDYFDHLESYERTQDLFANSTKTVQKETOAEY-----LLG- 102
Db 487 RIHDPFPVASSGLILSPYDNHL-----PEIKALLEKEKEKQKTSVEYHPFCKVMLLGN 540
QY 103 --LGR---GEYDTEYAAQARIGIKHIVLGI-----GPDVILGATRYTT 142
Db 541 HSSGKTTFLSQYDNTYTYQ-----KNTHVLSIHRSNPNNAIFYDFGGQDYHYGIYQAFET 595

RESULT 13
AAR07014
ID AAR07014 standard; protein; 593 AA.
XX AC AAR07014;
XX XX 24-OCT-2003 (revised)
XX DT 25-MAR-2003 (revised)
XX DT 17-JAN-1991 (first entry)
XX DE Protein G variant with three active sites.
XX KW Immunoglobulin.
XX OS Streptococcus sp; Lancefield Group G strain.
XX PH Key Location/Qualifiers
XX FT Active-site 303..372 /label= B1
XX FT Active-site 373..427 /label= B3
XX FT Active-site 443..497 /label= B2
XX PN US4956296-A.

```



```
Query Match      10.5%; Score 98; DB 2; Length 593;
Best Local Similarity 25.3%; Pred. No. 0.32;
Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;

QY 10 TADVRNGIDGHALADRIGLDEABIAWRLSFTGIDD-----DTMAALAAE----- 53
Db 41 TPIIRNGGE---LTNLLGNSETTLALRNESATADLTAAAVADTVAAAAAENAGAAAWEA 97
QY 54 ----QPLFEATADAL-----VTDFYDHLESYERTQDLFANSTKTVBOLKETOABYLLG 102
Db 98 AAAADALAKAKADALKEFNKYGVSDYKNL-----INNAKTVEGVKDLQAQVW-- 145
QY 103 LGRGEYDTEYAAQARARIGIKIHVDVILGLGPDVILGAYTRYTYTGLLDALADDVADRGEEAAA 162
Db 146 -----ESAKKARISEATD-----GLSDFLKSQTPA---EDTVK 175
QY 163 AVDELVAREL 172
Db 176 SIELAEAKVL 185
```

RESULT 15
AAR10005
ID AAR10005 standard; protein; 594 AA.
XX
AC AAR10005;
XX
DT 24-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 13-MAR-1991 (first entry)
XX
DE Streptococcus GX7805 protein G.
XX Immunoglobulins; Ig.
XX
OS Streptococcus sp; GX7805.
XX
FH Key Location/Qualifiers
FT Active-site 304..358
FT /label= Active Site B1
FT Active-site 374..428
FT /label= Active Site B3
FT Active-site 444..498
FT /label= Active Site B2
XX
PN US4977247-A.
XX
PD 11-DEC-1990.
XX
PF 19-MAY-1989; 89US-00354264.
XX
PR 14-FEB-1986; 86US-00829354.
PR 23-APR-1986; 86US-00854887.
PR 17-FEB-1987; 87WO-US000329.
PR 19-JUN-1987; 87US-0063959.
PR 20-JUN-1988; 88US-00209236.
XX
PA (GENEX) GENEX CORP.
XX
PI Fahnestock SR, Lee T, Wroble MH;
XX
DR WPI; 1991-006758/01.
DR N-PSDB; AAR10002.
XX
PT Immobilised protein G variants - used for detection, isolation and
PT purificn. immunoglobulin(s) and immunoglobulin fragments.
XX
PS Disclosure; Fig 9; 52pp; English.
XX
CC Protein G gene product may be modified allowing the variant to be
CC immobilised and exhibit different binding profiles. The bound protein is
CC useful in purification and detection of Igs and fragments. (Updated on 25

```
CC -MAR-2003 to correct PA field.) (Updated on 25-MAR-2003 to correct PI  
CC field.) (Updated on 24-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 594 AA;  
Query Match      10.5%; Score 98; DB 2; Length 594;  
Best Local Similarity 25.3%; Pred. No. 0.32;  
Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;  
QY 10 TADVRNGIDGHALADRIGLDEABIAWRLSFTGIDD-----DTMAALAAE----- 53  
Db 41 TPIIRNGGE---LTNLLGNSETTLALRNESATADLTAAAVADTVAAAAAENAGAAAWEA 97  
QY 54 ----QPLFEATADAL-----VTDFYDHLESYERTQDLFANSTKTVBOLKETOABYLLG 102  
Db 98 AAAADALAKAKADALKEFNKYGVSDYKNL-----INNAKTVEGVKDLQAQVW-- 145  
QY 103 LGRGEYDTEYAAQARARIGIKIHVDVILGLGPDVILGAYTRYTYTGLLDALADDVADRGEEAAA 162  
Db 146 -----ESAKKARISEATD-----GLSDFLKSQTPA---EDTVK 175  
QY 163 AVDELVAREL 172  
Db 176 SIELAEAKVL 185
```

Search completed: August 10, 2004, 15:26:16
Job time : 33.8547 secs

B/ANK

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	91.5	9.8	955	2	US-08-428-414A-3	Sequence 3, Appli
2	89	9.5	955	1	US-08-006-676B-1	Sequence 1, Appli
3	89	9.5	955	1	US-08-282-845-2	Sequence 2, Appli
4	89	9.5	955	5	PCT-US94-00324-1	Sequence 1, Appli
5	85.5	9.2	302	4	US-09-328-352-4846	Sequence 4846, Ap
6	83.5	8.9	319	4	US-09-489-039A-8872	Sequence 8872, Ap
7	82	8.8	542	4	US-09-252-991A-21396	Sequence 21396, Ap
8	81.5	8.7	1253	4	US-09-252-991A-30019	Sequence 30019, A
9	80.5	8.6	700	4	US-09-252-991A-19384	Sequence 19384, A
10	80	8.6	553	4	US-09-252-991A-32970	Sequence 32970, A
11	80	8.6	677	4	US-09-252-991A-18102	Sequence 18102, A
12	79	8.5	438	4	US-09-540-236-2595	Sequence 2595, Ap
13	79	8.5	878	3	US-08-941-936-2	Sequence 2, Appli
14	78.5	8.4	733	4	US-09-328-352-5599	Sequence 5599, Ap
15	78.5	8.4	755	5	PCT-US93-07923-3	Sequence 3, Appli
16	78.5	8.4	759	5	PCT-US93-07923-2	Sequence 2, Appli
17	78.5	8.4	766	1	US-08-230-491A-3	Sequence 3, Appli
18	78.5	8.4	766	1	US-08-619-280A-3	Sequence 3, Appli
19	78.5	8.4	766	2	US-08-940-391-3	Sequence 3, Appli
20	78.5	8.4	766	4	US-09-794-236-1	Sequence 1, Appli
21	78.5	8.4	766	4	US-10-003-593-6	Sequence 6, Appli
22	78	8.4	443	4	US-09-328-352-6943	Sequence 6943, Ap
23	78	8.4	510	4	US-09-489-039A-11123	Sequence 11123, A
24	77	8.3	297	4	US-09-252-991A-17492	Sequence 17492, A
25	77	8.3	320	4	US-09-252-991A-21006	Sequence 21006, A
26	76.5	8.2	579	4	US-09-252-991A-18063	Sequence 18063, A
27	76	8.1	531	4	US-08-976-063E-34	Sequence 34, Appli

OPERATING SYSTEM: Apple System 7.1
 SOFTWARE: Microsoft Word, version 5.1a
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/00324
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/006,676
 FILING DATE: 15-JAN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Perkins, Patricia Anne
 REGISTRATION NUMBER: 34,693
 REFERENCE/DOCKET NUMBER: 5004-WO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 587-0430
 TELEFAX: (206) 233-0644
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 955 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US94-00324-1

Query Match 9.5%; Score 89; DB 5; Length 955;
 Best Local Similarity 29.3%; Pred. No. 0.3;
 Matches 49; Conservative 19; Mismatches 69; Indels 30; Gaps 7;
 QY 44 DDTMAALAEQPLFEATADALVT---DFYDHLSEYVETQDLFANSTKTVQELKETOAEYL 100
 DB 546 ESTVAQLREQERREVALDQTHQKLEALESSERTA---AERDQLQLQELQSE-R 601
 QY 101 LGLGRGEYDTE-YAAQARIGKIHDLVGLGPDVVLGA---YTRYVTGL----- 145
 DB 602 TQLSQVVTDRRLTRDLRLQIQEYGETELARVDVALCAQEMEARHAAVFHLQTLLELAT 661
 QY 146 ---DALADVADRGEAAAVDELV-----ARFLPMLKLLTFDQOI 184
 DB 662 EWEDALRERALARDEMAAELDAAASTQNAESACELTSLQQL 708

RESULT 5
 US-09-328-352-4846
 ; Sequence 4846, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 4846
 ; LENGTH: 302
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-4846

Query Match 9.2%; Score 85.5; DB 4; Length 302;
 Best Local Similarity 29.4%; Pred. No. 0.14;
 Matches 37; Conservative 23; Mismatches 41; Indels 25; Gaps 7;
 QY 25 RIGL---DEAEIATWLSFTGIDDDTMAALAEQPLFEATADALVTDFYDHLSEYVETQDL 81
 DB 179 RIGLSNVDSAEILKSLP-QGPHNNTLVYL---DPEYAKGQDLXENFYNHQDHVEIMRAL 234
 QY 82 FANSTK-----TVEQLKETQAEYLLGLGRGEYDTEYAAQARIGK-----INDVLGLG 129
 DB 235 KSSSIKWIVSYDNDVAIRELYKDFRV---LEYSLQTAQOKKIGEEVWFISNDV--LI 288
 QY 130 PDVYIG 135

Db 289 PNVRIG 294

RESULT 6
 US-09-489-039A-8872
 ; Sequence 8872, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 8872
 ; LENGTH: 319
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-8872

Query Match 8.9%; Score 83.5; DB 4; Length 319;
 Best Local Similarity 25.4%; Pred. No. 0.26;
 Matches 51; Conservative 17; Mismatches 72; Indels 61; Gaps 9;
 QY 14 RINGIDGHALADRIGLDEAEIA-WRLSFTGIDDDTMAALAEQPLF-----EATADALV 65
 DB 120 RNRIDGVVLFSGTIDEAMLAPWR-----DTLVLMARDAPGFASVCYDDEGAILLM 171
 QY 66 TDFYDH-----LESYERTQDLFANSTKTVQELKETOAEYLLGLGRGEYDTE 111
 DB 172 QRLYDRGHRHSFLGVPHSDVTTGERRHLAYLAFCK-HRLTPTAALPGLGMKQG-YDTV 229
 QY 112 YAAQARIGKI---HDVLGLGPDVVLGAYTRYVTGLLDAL----- 148
 DB 230 ASVLTAETSALVCATDTTALGASKYLOOQGR-----DALQLASVGSVTPLMKFLHPEILT 283
 QY 149 ADDVVADRGEAAAVDELVA 169
 DB 284 VDPGVAESGRRAARQLIEQIA 304

RESULT 7
 US-09-252-991A-21396
 ; Sequence 21396, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 21396
 ; LENGTH: 542
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-21396

Query Match 8.8%; Score 82; DB 4; Length 542;
 Best Local Similarity 25.3%; Pred. No. 0.84;
 Matches 49; Conservative 19; Mismatches 66; Indels 60; Gaps 8;
 QY 3 NDNDTLVTADVRNGIDGHALADRIGLDEAEIATWLSFTGIDDDTMAALAEQPLFEATAD 62
 DB 149 NENDTVVTDEIRFG-----DNDTLAALVAN--LVEADLL 180


```

: Patent No. 6551795
:
: GENERAL INFORMATION
:
: APPLICANT: Marc J. Rubenfield et al.
:
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
:
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
:
: FILE REFERENCE: 107196.136
:
: CURRENT APPLICATION NUMBER: US/09/252,991A
:
: CURRENT FILING DATE: 1999-02-18
:
: PRIOR APPLICATION NUMBER: US 60/074,788
:
: PRIOR FILING DATE: 1998-02-18
:
: PRIOR APPLICATION NUMBER: US 60/094,190
:
: PRIOR FILING DATE: 1998-07-27
:
: NUMBER OF SEQ ID NOS: 33142
:
: SEQ ID NO 18102
:
: LENGTH: 677
:
: TYPE: PRT
:
: ORGANISM: Pseudomonas aeruginosa
:
: US-09-252-991A-18102

```

Query Match	8.6%;	Score 80;	DB 4;	Length 677;
Best Local Similarity	26.6%;	Pred. No. 2;		
Matches	45;	Conservative	18;	Mismatches 64; Indels 42; Gaps 10;

QY	6	DTLVTDVRNGIDGHALADRGILDEARIWLSFTGDDDTMAALAAEQPIFEATADALV	65
		: :	
Db	245	DKLUSGDPTPMVADEVPRDLNAR-----SWR-----AFIGDFPLSRDREALI	288
		: :	
QY	66	TDFYDHLESYERTQDLFANSTKTKVEQ-----LKETAQAEYLL-GLGRGEYDTEYAQAQRARI	119
		:	
Db	289	A-----LYESPRDYLAG--KSVEEEKETYLAKTSYRDYLLKNVLSETSVKKTFQ-----	334
QY	120	GKHHDVLGGPDVYLGAYTYYTGL---LDALADDVADRGEERAAAAVDE	166
		: : : :	
Db	335	GRSNDFSALGADA-LPAADAYAAGFGPGFDALG---LPQPSECAAEWDE	379

```

RESULT 12
US-09-540-236-2595
; Sequence 2595, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2595
; LENGTH: 438
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2595

```

Query Match	8.5%; Score 79; DB 4; Length 438;
Best Local Similarity	23.8%; Pred. No. 1.4;
Matches	53; Conservative 73; Mismatches 73; Indels 70; Gaps 10
QY	17 IDGHALADRIGLDAAETAWRLSFTGIDDDTMAALAAEQPLFEATDALVTDFDHLSYE 76
dDb	: : : : : : : : :
QY	19 IDPQLANALPHFE-YDGTILHFDGINHTLKEIYG-TPUIAYSQRTL-----LENQ 69
dDb	: : : : : : : : :
QY	77 RTQDLF-----ANST-----KT-----VEQLKETQAEYLIG 102
dDb	: : : : : : : : :
QY	70 SYTFADFDAIDHQICVAVKANSNALIKTLAKAGAFDIVSGELARVLQVADAKKVYVG 129
dDb	: : : : : : : : :
QY	103 LGRGEYDEVAQR-----ARIKGTHDVLG-----LGPDVILGAYTRY 141
dDb	: : : : : : : : :
QY	130 VGTKADDIEALXADIHCNFVESISELDTINVAGRQNQIARISURINPDVDKATHPVIS 189
dDb	: : : : : : : : :
QY	142 TGLLDALADVVDARGEEAAAADVELVARFLPKLLTFTDQOI 184
dDb	: : : : : : : : :
QY	190 TG-----MKDNKFGISHQQAAY--VVAHHLPNKLIVGDCHI 226
dDb	: : : : : : : : :

```

RESULT 13
US-08-941-936-2
; Sequence 2, Application US/08941936
; Patent No. 6054305
; GENERAL INFORMATION:
; APPLICANT: Tatsumi, Hiroki
; APPLICANT: Eisaki, Naoki
; APPLICANT: Horiuchi, Tatsuo
; APPLICANT: Nagahara, Ayumu
; TITLE OF INVENTION: Pyruvate Orthophosphate Dikinase Gene,
; TITLE OF INVENTION: Recombinant DNA, and Process For Producing Pyruvate
; TITLE OF INVENTION: Orthophosphate Dikinase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/941.936
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: HIRAKI-03009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-705-8410
; TELEFAX: 415-397-8338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-941-936-2

```

Query Match	8.5%	Score	79	DB	3	Length	878
Best Local Similarity	25.3%	Pred. No.	3.9				
Matches	49	Conservative	26	Mismatches	67	Indels	52
						Gaps	10
QY	22	LADRIGLDEAEIAWRLSFTGID-----DDTMAALAAEQ-----	PLPEATADA--L	VTD	67		
Ddb	329	LOTVGVKRTAEAFRIATQIVDGLTMDAEAVARVTGQDLAQIMPFERFAATADARLLTTG					
QY	68	FY-----DHLESYVERTODLFANSTKTVQEQLKETQAEYLLGL--GRGEY-----DTEY	112				
Ddb	389	MNASPGAAVGKAVFSSERAVELAGQGEAVILVRRETNPDLAGMTAARGVLTSRGKGTSH	448				
QY	113	AAQPARIGKTHDVLGPGDVLCA-----YTRYIT--GLLDALADDDVWDRGEEAAA	163				
Ddb	449	AAVVAR-----GMKCTCVCGAEELEVPDHARRFTAPGGIVVNGEYISIDGSTGAVY	500				
QY	164	VDEL-----VARF	171				
Ddb	501	LGVEPVTASPVARY	514				

RESULT 14
US-09-328-352-5599
; Sequence 5599, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.

```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5599
; LENGTH: 733
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5599

Query Match      8.4%; Score 78.5; DB 4; Length 733;
Best Local Similarity 21.6%; Pred. No. 3.4; 74; Indels 83; Gaps 10;
Matches 52; Conservative 32; Mismatches 74; Indels 83; Gaps 10;

QY 9 VTADVNRGIDGHALADR-----IGLDEAEIAWRLSFTGIDDD-----TMAALA 51
Db 223 IPAAVDNITSSNSVAERGKLTGILG-----LNYQGYDKHEHITQVLNAILVTYGAQN 273
QY 52 AEQPLFEAT-----ADALVTDYDHLHESYERTQDLPANSTKTYE-----90
Db 274 VERRSAESAQTILKFLDEQLPDLKKQLDDAEROFNKFROQYNTVDVTKESELYLTOSITLE 333
QY 91 -----OLKETQAEVLLGLGRGEYDTEYAAOR-----ARIGKIHDLVGLGPDV---Y 133
Db 334 TKKAELEQQAEM-----AAKYTAEPHPAMREINGQITAINKQIGELNSTLKQLPDVQRY 388
QY 134 LGAY-----TRYVTGLLDALADDVVADRGE-----EAAAADVELVARFLPMIKLIT 179
Db 389 LQLYREVEVKTQLYTALLNSYQQLRIAKAGIEGNVRIVDTAVEPVEPIKPKKLQVLISL 448
QY 180 F 180
Db 449 F 449

RESULT 15
PCT-US93-07923-3
; Sequence 3, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schlossman, Stuart F.
; APPLICANT: Tanaka, Toshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: Wordperfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07923
; FILING DATE: 19930819
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/832,211
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070

```

```

; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 755
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
PCT-US93-07923-3

Query Match      8.4%; Score 78.5; DB 5; Length 755;
Best Local Similarity 25.8%; Pred. No. 3.5;
Matches 47; Conservative 25; Mismatches 73; Indels 37; Gaps 8;

QY 2 SNDNDTLVTADVNR-NGIDG-----HALADRIGL-----DRAEIAWRLSFTGIDDD-----45
Db 558 STENIIIVASFDRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQFSKMGFVNDKRIAWG 617
QY 46 -----TMAALAAEQPLFE-----ATA-----DALVTDFYDHLHESYERTQDLPANSTK 87
Db 618 WSYGGVVTSMVLGSGGVFKCGIAVAPVSRWEYDVSVYTERVMGLPTPEDNLDHYRNV 677
QY 88 TVEQLKETQAEVLLGLGRGEYDTEYAAQPARIGKIHDLVGLGPDVYVGLGAYTRYVTGLLDA 147
Db 678 MSRAENFKQVEYLLIHGTAD-DNVHFQSSAQISKA--LVDVGVDFQAMWYTTDEDHGIASS 734
QY 148 LA 149
Db 735 TA 736

Search completed: August 10, 2004, 15:30:20
Job time : 9.35593 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:28:46 ; Search time 26.7312 Seconds
(without alignments)
2159.184 Million cell updates/sec

Title: US-09-455-978B-77

Perfect score: 933

Sequence: 1 MSNDNDTLVTADVRNGIDGH.....DELVARFLPMLKLLTFDDQI 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	99	10.6	448	12	US-10-460-524-5
2	94.5	10.1	883	15	Sequence 5, Appli
3	94	10.1	218	14	Sequence 18563, A
4	89.5	9.6	400	14	Sequence 9281, Ap
5	88	9.4	505	12	Sequence 13480, A
6	87	9.3	496	12	Sequence 66025, A
7	86	9.2	362	16	Sequence 6364, A
8	85	9.1	2703	12	Sequence 137315, A
9	83.5	8.9	258	14	Sequence 66108, A
10	83.5	8.9	315	12	Sequence 11335, A
11	83.5	8.9	367	12	Sequence 59961, A
12	83.5	8.9	368	12	Sequence 72593, A
13	83.5	8.9	1254	12	Sequence 48921, A
14	83	8.9	384	12	Sequence 53778, A
15	82.5	8.8	403	12	Sequence 48636, A
					Sequence 61978, A

16	82.5	8.8	695	12	US-10-282-122A-69832	Sequence 69832, A
17	82.5	8.8	765	15	US-10-369-493-13529	Sequence 13529, A
18	82.5	8.8	831	9	US-09-738-626-5468	Sequence 5468, Ap
19	82.5	8.8	928	12	US-10-282-122A-61748	Sequence 61748, A
20	82.5	8.8	939	14	US-10-156-761-12605	Sequence 12605, A
21	82	8.8	118	12	US-10-425-114-63899	Sequence 63899, A
22	82	8.8	329	12	US-10-289-456-16	Sequence 16, Appl
23	82	8.8	329	12	US-10-622-064-28	Sequence 16, Appl
24	82	8.8	329	14	US-10-243-739-16	Sequence 28, Appl
25	82	8.8	329	14	US-10-244-065-16	Sequence 16, Appl
26	82	8.8	329	14	US-10-289-454-16	Sequence 16, Appl
27	82	8.8	329	15	US-10-346-190-16	Sequence 16, Appl
28	82	8.8	329	15	US-10-465-811-7	Sequence 7, Appl
29	82	8.8	329	16	US-10-622-124-10	Sequence 10, Appl
30	82	8.8	329	16	US-10-622-087-10	Sequence 10, Appl
31	82	8.8	330	14	US-10-050-902-254	Sequence 254, App
32	82	8.8	330	14	US-10-050-898-254	Sequence 254, App
33	82	8.8	901	15	US-10-369-493-11290	Sequence 8893, Ap
34	82	8.8	1333	12	US-10-282-122A-55546	Sequence 18812, A
35	82	8.8	2799	12	US-10-282-122A-65564	Sequence 55546, A
36	81	8.7	1178	14	US-10-128-714-8240	Sequence 8240, A
37	80.5	8.6	546	15	US-10-369-493-8893	Sequence 8893, Ap
38	80.5	8.6	1173	15	US-10-369-493-11290	Sequence 11290, A
39	80.5	8.6	2132	12	US-10-424-599-161362	Sequence 161362, A
40	80.5	8.6	2151	12	US-10-424-599-161366	Sequence 161366, A
41	80.5	8.6	2221	12	US-10-424-599-161365	Sequence 161365, A
42	80	8.6	811	16	US-10-437-963-171424	Sequence 171424, A
43	80	8.6	954	12	US-10-282-122A-69774	Sequence 69774, A
44	79.5	8.5	429	12	US-10-282-122A-62133	Sequence 62133, A
45	79.5	8.5	1015	16	US-10-437-963-184684	Sequence 184684, A

ALIGNMENTS

RESULT 1

US-10-460-524-5
; Sequence 5, Application US/10460524
; Publication No. US20040029781A1
; GENERAL INFORMATION:
; APPLICANT: Hernan, Ronald A
; APPLICANT: Mehig, Richard J
; APPLICANT: Brockie, Ian
; APPLICANT: Jenkins, Elizabeth
; TITLE OF INVENTION: Affinity Peptides and Method for Purification of Recombinant Proteins
; FILE REFERENCE: SGM 7047.1
; CURRENT APPLICATION NUMBER: US/10/460,524
; CURRENT FILING DATE: 2003-06-12
; PRIOR APPLICATION NUMBER: US 60/388,059
; PRIOR FILING DATE: 2002-06-12
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Streptococcus
US-10-460-524-5

Query Match	10.6%	Score 99;	DB 12;	Length 448;
Best Local Similarity	25.3%	Pred. No. 0.087;	Mismatches 45;	Indels 72; Gaps 8;
Matches	48;	Conservative 25;		
Qy	10	TADVRNGIDGHADRLGDEABIAWRLSPTGIDD-----DTMAALAAE-----	53	
Db	41	TPRIIRNGE---LTNLLGNSETTLALRNEESATADLTAAAVADTVAAANAGAAWEA	97	
Qy	54	-----QPLFETADAL-----VTFDYHLESYERTQD;FANSTKTVEQKETQAEYLIG	102	
Db	98	AAAAADALAKAKADAKAFKFNKYGVSDYKXNL-----INNAKTVEGKDKIQAQVW--	145	
Qy	103	LGRGEVDTEVAQORARIGKIHVDLGLGPDVYLCAITRYTGLLDALADDDVADRGEEAA	162	
Db	146	-----ESAKKARISATD-----GLSDFLKSTQTPA---EDTVK	175	

```
QY 163 AVDELVARFL 172
Db 176 SIELAEAKVL 185

RESULT 2
US-10-369-493-18563
; Sequence 18563, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18563
; LENGTH: 883
; TYPE: PRT
; ORGANISM: Halobacterium sp. NRC-1
US-10-369-493-18563

Query Match 10.1%; Score 94.5; DB 15; Length 883;
Best Local Similarity 28.3%; Pred. No. 0.67; Mismatches 23; Indels 43; Gaps 10;
Matches 51; Conservative 23;

QY 12 DVNRGIDGH--ALADRIGLDEAEI--AWRLSFTGIDDDTMAALAAEQPLFEA-----TAD 62
Db 180 DVKSNEGGQLDLAQIA--DKEADPHRLASHNTALAEVTDIEHFEAREQARQTRD 237
QY 63 ALVTDFYHLESYERTQDLFANSTKTVEQKLTQAEYLLGLGRGEYDTEYAAQRA----- 117
Db 238 ----DAADVLYEESRTALADVEETIADVREAVAE-----AERERTLADRVSDHRE 286
QY 118 RIGKIHV-----LGL-GPDVYLGAYTRYTGLLDALAD--DVVADRGEAAAVDEL 167
Db 287 RASLDDEAAALADLGLDDPDPAEDASAER-----DAVADQREAVAEVRVAPAVSRL 340

RESULT 3
US-10-156-761-9281
; Sequence 9281, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9281
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9281

Query Match 10.1%; Score 94; DB 14; Length 218;
Best Local Similarity 28.7%; Pred. No. 0.11; Mismatches 22; Indels 22; Gaps 8;
Matches 47; Conservative 22;

QY 21 ALADRIGLDEAEI--AWRLSF-TGIDDDTMAALAAEQPLFEATADALVTDYDHLSEYER 77
Db 25 ALADRLGVAHAEVDFVGRSLSVASGVPEPVVKALLSGRPAGEPDLQA---RFLQRLDLLRR 81
QY 78 TQDLFANSTKTVEQ-----LKETOAEYLLGLGRGEYDTEYAAQRAKIGIHD--VLGL 128
Db 82 TR-LKPNRRRYTQOEIADGAGMSRQQAGALIN-GDRRPTMEHCDAIQRFFRVHAGFLTA 139
QY 129 GPDVYLGAYTRYTGLLDALADDVADVGRGEAAAVDELVARFL 172
Db 140 DPEALAGTLQRSEQLLOOL-----ADRERAAAMAVDDPFLERLL 178

RESULT 4
US-10-156-761-13480
; Sequence 13480, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13480
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13480

Query Match 9.6%; Score 89.5; DB 14; Length 400;
Best Local Similarity 26.7%; Pred. No. 0.78; Mismatches 66; Indels 17; Gaps 5;
Matches 36; Conservative 16;

QY 42 IDDDTMAALAAEQPLFEATADALVTDYDHLSEY--ERTQDLFANSTKTVEQKLTQAEY 99
Db 2 LSEQSNATVTRATLPVGAAGVGEITAFYDRLFAARPELLRDLFNRG---NQAAGTORQA 57
QY 100 LLG--LGRGEYDTEYAAQRA-----RIGKIHVILGLGPDVYLGAYTRYTGLLDALADD 152
Db 58 LAGSIAAFATYLVHEHDERPDMALDRIAHKHASLGIAPGQYAVVHBLFAAIAEVLGDV 117
QY 153 VADRGEAAAVDEL 167
Db 118 T-----PEVAAAMDEV 128

RESULT 5
US-10-282-122A-66025
; Sequence 66025, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
```

```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66025
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-282-122A-66025

Query Match
Best Local Similarity 9.4%; Score 88; DB 12; Length 505;
Matches 46; Conservative 24; Mismatches 69; Indels 40; Gaps 7;

QY 6 DTLVTADVNRGIDGHALADRIGLDEAEIAWRLSP---TGIDDDTWAALAAEQPLF----- 57
Db 325 DQALAIAGRGQNVRLASDLTG-----WQINIMTSAEADERNAEDAAIRRLFMHNLN 377
QY 58 --EATADALVTDFDHLR--SVERTQDLFANSTKTVEQLKETOAEYLLGLGRGEYDTBYA 113
Db 378 VDEETADVIVQEGFATLEEVAVPAAELLA-----IEGFDEIVDMLRNRARDAILTWAI 432
QY 114 AORARIGKIHDLV---GLGPDVYILGAYTRYTGLLDALADDVVDRCGEAAAVDELV 168
Db 433 AAEEKLGEVSDMRNLEGIDAD-----MLRSLAEGITTRDDLAELAVDELI 479

RESULT 6
US-10-282-122A-65364
; Sequence 65364, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
```

```
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65364
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-282-122A-65364

Query Match
Best Local Similarity 9.3%; Score 87; DB 12; Length 496;
Matches 46; Conservative 24; Mismatches 69; Indels 40; Gaps 7;

QY 6 DTLVTADVNRGIDGHALADRIGLDEAEIAWRLSP---TGIDDDTWAALAAEQPLF----- 57
Db 316 DRALAIAGRGQNVRLASDLTG-----WQINIMTSAEADERNAEDAAIRRLFMHNLN 368
QY 58 --EATADALVTDFDHLR--SVERTQDLFANSTKTVEQLKETOAEYLLGLGRGEYDTBYA 113
Db 369 VDEETADVIVQEGFATLEEVAVPAAELLA-----IEGFDEIVDMLRNRARDAILTWAI 423
QY 114 AORARIGKIHDLV---GLGPDVYILGAYTRYTGLLDALADDVVDRCGEAAAVDELV 168
Db 424 AAEEKLGEVSDMRNLEGVDADMLL-----SLAEGITTRDDLAELAVDELI 470

RESULT 7
US-10-437-963-137315
; Sequence 137315, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 137315
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Oryza sativa
```

```
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_3880C.1.pcp
US-10-437-963-137315

Query Match      9.2%; Score 86; DB 16; Length 362;
Best Local Similarity 26.4%; Pred No. 1.6;
Matches 51; Conservative 20; Mismatches 78; Indels 44; Gaps 9;

QY 8 LVTADVRNGIDGHADLADRIGLDEAEIAWR-----LSFT--GIDDDTMAALAAPOQLFEATA 61
Db  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26 ILADESTGTIGKRLAS-IGVENVEENRRALRELLFTAPGALDCLSGVILFEETLYQSTR 84
QY 62 DALVTDFVDHLESY-----ERTQDLFANSTYTVQLKETQAEYLLGLGRGEVDTEY 112
Db  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 DG--TPFVDVLAAGVLAGIKVDKGTVELAGTDRETTQGH-----GLGRCRRY 133
QY 113 AAOARIGIKIHVDVLGLG-----PDVILGAVTRY-----YTGLLDALADDVVAD-- 155
Db  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 YAAGARFAKRAVLISIGRASSRPSQLAVDANAQGLARVAIIQENGLVPIVEPEILVDGE 193
QY 156 RGEBAAAAVDELV 168
Db  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 HGIEACAETERV 206

RESULT 8
US-10-282-122A-66108
; Sequence 66108, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66108
; LENGTH: 2703
; TYPE: PRT
; ORGANISM: Neisseria meningitidis

US-10-282-122A-66108
Query Match      9.1%; Score 85; DB 12; Length 2703;
Best Local Similarity 23.5%; Pred. No. 32;
Matches 52; Conservative 26; Mismatches 71; Indels 72; Gaps 9;

QY 3 NDNDDTLVTADVRNGIDGHAL-----ADRIGLDEAEIAWRLSFTGIDDDT----- 46
Db  2153 SDGUS-KNSTRTSGVNTNHIITDEAGQLARTGRTAKETEARI-YTGIDTETADQHSGLH 2210
QY 47 -----MAALAAPOQLFE-----ATAALVTDFVDHLESYERTQDLFANSTYTVQ 91
Db  2211 KNSFKDAVAKBINLQREVTKEFGRNAQAQAAVADKLNGTQSVRYQEA---RTLLEAE 2267
QY 92 LKETOAEYLLGLGRGEYDEYAAQPARIGKIHDLV-----GLGPDVVLGAYT 138
Db  2268 LQNTDSE-----AEKAAFRASLGQVNAVYLAENQSRDYDTWKEGGIGRSILHGAAG 2316
QY 139 RYVTGLLDAL-----ADDVADRGEEAAAVDEL 167
Db  2317 GLTTGSLGILLAGGTSLAAPYLDKAAENLGPAGKAAVNAL 2357

RESULT 9
US-10-156-761-11335
; Sequence 11335, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11335
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11335

Query Match      8.9%; Score 83.5; DB 14; Length 258;
Best Local Similarity 24.1%; Pred. No. 1.9;
Matches 49; Conservative 21; Mismatches 78; Indels 55; Gaps 9;

QY 8 LVTADVRNGIDGHADLADRIGLDEAEIAWRLSFTGIDDDTMAALAAPOQLFEATADALVTD 67
Db  33 VVADVLDLDD-QGEALAKEIGARYVHLDD-----VGREDDWQAAV-----TVAKD 74
QY 68 FVDHLESY-----RTQDLFANSTYTVQQLKE-TQAEYLLGLGRGEYDEYAAQPARIG 120
Db  75 AYCHIDGLVNNAGILRFNDLVGTPLAABFQQIVQVNVGVFLGIKTVAPIE-AAGGGTIV 133
QY 121 KIHVDVLGLGPDVVLGAYT-----RYVTGLLD-ALADDVVA 154
Db  134 NTASYAGLTGMAYVGYATTKHAIUGLTVRALEAAKIRVNAVCPGADITAMSPSQL 193
QY 155 DRG---EAAAAVDELVARFLPM 174
Db  194 DPGADPEETARALSSELYGRVLPL 216

RESULT 10
US-10-282-122A-59961
```



```

; Sequence 59961, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59961
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-59961

Query Match      8.9%; Score 83.5; DB 12; Length 315;
Best Local Similarity 25.4%; Pred. No. 2.4;
Matches 51; Conservative 17; Mismatches 72; Indels 61; Gaps 9;

QY 14 RRGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLF-----ETADALV 65
DB 116 RRNDGVVLGFTGIDEAMLAAPWR-----DTLVLMARDAPGFASVCYDDEGAITLM 167
QY 66 TDFYDH-----LESYRTQDLFANSTKTVEQLKETAQAEYLLGLGRGEYDTE 111
DB 168 QRLYDRGHRHSFLUGVPHSDVTTCGERRHAYLAFCKK-HRLTPTAALPGLGMKQG-YDVT 225
QY 112 YAAQBARIGKI-----HVLGIGPQVYLVGAYTRYTGTLLDAL----- 148
DB 226 ASVLTAFTSALVCATDTLTALGASKYLQOQGR-----DALQLASVSGSTPLMKFLHPILT 279
QY 149 ADDVADRGEEAAAVDELVA 169
DB 280 VDPGYAESGRRAARQLIEQIA 300

RESULT 11
US-10-425-114-72593
; Sequence 72593, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48921
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: CC-ZMP0148032B07_FLI.pep
US-10-425-114-48921

Query Match      8.9%; Score 83.5; DB 12; Length 368;
Best Local Similarity 25.3%; Pred. No. 3;
Matches 47; Conservative 24; Mismatches 68; Indels 47; Gaps 10;

QY 16 GIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFETADALVTDYFDHLESY 75
DB 152 GVDGR-----RVGLDEEAPA-----DADAVAAQAHRDHLDD--VEALLVELVDLGH- 196
QY 76 ERTQDLFANSTKTV-----EQLKETAQAEYLLGLGR-----GEYDTEVAAG-----R 116
DB 197 ---RDVRAQVGVVLAUGHVAPPQGHVRLGPARQLHRVAGAHRDDVRAHHVRAILLIQ 253
QY 117 ARIGKIHVDVLGIGPDV-----VLGAYTRY---VTGLDALDDVVADRGEEAAAVDELVA 169
DB 254 LRLDLLDHVVALHGDVGAGLLGLVRRHVQDQGRVAICAVVEDEAEERAGGAG---V 310
QY 170 RFLPWL 175
DB 311 SFEPLL 316

RESULT 12
US-10-425-114-48921
; Sequence 48921, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48921
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: CC-ZMP0148032B07_FLI.pep
US-10-425-114-48921

Query Match      8.9%; Score 83.5; DB 12; Length 368;
Best Local Similarity 25.3%; Pred. No. 3;
Matches 47; Conservative 24; Mismatches 68; Indels 47; Gaps 10;

QY 16 GIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFETADALVTDYFDHLESY 75

```

```

; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 72593
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: CC-ZMP014820017B02_FLI.pep
US-10-425-114-72593

Query Match      8.9%; Score 83.5; DB 12; Length 367;
Best Local Similarity 25.3%; Pred. No. 3;
Matches 47; Conservative 24; Mismatches 68; Indels 47; Gaps 10;

QY 16 GIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFETADALVTDYFDHLESY 75
DB 152 GVDGR-----RVGLDEEAPA-----DADAVAAQAHRDHLDD--VEALLVELVDLGH- 196
QY 76 ERTQDLFANSTKTV-----EQLKETAQAEYLLGLGR-----GEYDTEVAAG-----R 116
DB 197 ---RDVRAQVGVVLAUGHVAPPQGHVRLGPARQLHRVAGAHRDDVRAHHVRAILLIQ 253
QY 117 ARIGKIHVDVLGIGPDV-----VLGAYTRY---VTGLDALDDVVADRGEEAAAVDELVA 169
DB 254 LRLDLLDHVVALHGDVGAGLLGLVRRHVQDQGRVAICAVVEDEAEERAGGAG---V 310
QY 170 RFLPWL 175
DB 311 SFEPLL 316

```

Db 152 GVDR-----RVGLDEEAPA-----DADAVAAQHRDHLDD--VEALLVELVDELGH- 196
QY 76 ERTQDLFANSTKTV-----EQKETOAEYLLGLGR-----GEYDTEYAAQ-----R 116
Db 197 ---RDVRAQVGVGLAVGHERVAPPQGHVGLGPARGQLHRVAGAHRDDVRAHHVRLILQ 253
QY 117 ARIGKIHVDVGLGPDV-----VLGATRY--YTGLLDALDDVVDVDRGEEAAAADVELVA 169
Db 254 LRLDLHDVHALHGVGLGLLGLVLRHRVDQDGRVAAICBAVVEDEAEAGGGAG-----V 310
QY 170 RFLPML 175
Db 311 SFEPLL 316

RESULT 13

US-10-282-122A-53778

; Sequence 53778, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELIPRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 53778

; LENGTH: 1254

; TYPE: PRT

; ORGANISM: *Corynebacterium diptheriae*

US-10-282-122A-53778

Query Match 8.9%; Score 83.5; DB 12; Length 1254;

Best Local Similarity 25.0%; Pred. No. 16;

Matches 52; Conservative 23; Mismatches 82; Indels 51; Gaps 10;

QY 4 DNDTLV-----TAVRNSIDGHALADIGLDE-----REIAWRLSFTG 41

Db 658 DNLNLVNHVALFDPNEDLRNAADGAGYDDITLDEFVNTAPQMRFLGLVKAWELR-DG 716

QY 42 IDDDTMAALAAEQFLFATADALVTDYDHLSEYERTQDLFANSTKTVQKETAQAEVLL 101
Db 717 AGPQVQA--AADSALKGTGTPAL--DEFVNGDGYEKAR--YLDQVQQAAYELTDT----- 764
QY 102 GLGRGEYDTEYAAQARARIGKIHVDVGLGPDVVLGATRYTGLLDALAD-----DVVAD 155
Db 765 ---GGPEVQTAEEAAVTG---DRQQLDEFVSIQGYRR---AIIIDSORDAHNAEINALIS 814
QY 156 RGEAAAAAVDELVARFLPMLKLLTFFDQ 183
Db 815 AQONAAELASQEAANAQEAAYTRATGDAQ 842

RESULT 14

US-10-425-114-48636

; Sequence 48636, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E.

; APPLICANT: Tabaska, Jack E.

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 48636

; LENGTH: 384

; TYPE: PRT

; ORGANISM: *Zea mays*

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB31116-002-E8_FLI.pep

US-10-425-114-48636

Query Match 8.9%; Score 83; DB 12; Length 384;

Best Local Similarity 25.6%; Pred. No. 3.6;

Matches 31; Conservative 25; Mismatches 43; Indels 22; Gaps 5;

QY 58 EATADALVTDYDHLSEYERTQDLFANSTKTVQKETAQAEVLLGLGRGEYD 109

Db 267 KANSAPITAYRYHSLCSYMGDDMF--SSDLSEDLQRQLGHMSITQCQVIFSMG-DEYV 324

QY 110 TEYAAQARARIGKIHVDVGLGPDVVLGATRYTGLLDALDDVVDVDRGEEAAAADVELVA 169

Db 325 PEYVDKRALDLRLCALGGAKEVEI-----EWGNHALSNRQEAQVVAIVDFVK 372

QY 170 R 170

Db 373 R 373

RESULT 15

US-10-282-122A-61978

; Sequence 61978, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

```
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61978
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Mycobacterium avium
US-10-282-122A-61978

Query Match      8.8%; Score 82.5; DB 12; Length 403;
Best Local Similarity 25.4%; Pred. No. 4.4;
Matches 36; Conservative 20; Mismatches 65; Indels 21; Gaps 6;

QY 42 IDDDTMAALAAQPLFEATADALVTDYDHL--ESYERTQDLFANSTKTVEQLKETQAEY 99
Db 25 ISDSVLDALLAQDPRSRAVAVETLVTGQVHVGEVTTTAKAFADITNTVR-----ER 77

QY 100 LLGLGRGEYDTEYAAQRAKIGKIHVILGLG---PDVYLGAYTRYT---GLLDALADDVV 153
Db 78 ILDIGYDSSDKGFGDASCQVN-----IGGAQSPDIAQGVDTAHETRVGAADPLDAQGA 132

QY 154 ADRGEAAAAAVALDELVARF-LPM 174
Db 133 GDQGLMFGYAINDTPERMPLPI 154
```

Search completed: August 10, 2004, 15:42:28
Job time : 27.7312 secs

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:22:45 ; Search time 8.01937 Seconds
(without alignments)
2207.061 Million cell updates/sec

Title: US-09-455-978b-77
Perfect score: 933
Sequence: 1 MSNDNDTLVTADVRNGIDGH.....DELVARFLPMLKLLTFDQOI 184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	933	100.0	489	2 E84304	Htr10 transducer
2	933	100.0	489	2 T44978	transducer protein
3	148	15.9	439	2 A83713	methyl-accepting c
4	142	15.2	432	2 C69832	methyl-accepting c
5	132	14.1	537	2 C87302	methyl-accepting c
6	117.5	12.6	499	2 A97485	methyl-accepting c
7	117.5	12.6	499	2 A82703	methyl-accepting c
8	105.5	11.3	555	2 D87536	methyl-accepting c
9	99	10.5	448	2 A24436	IgG-binding protei
10	98	10.5	593	2 S00128	protein G precursor
11	97	10.4	218	2 T35174	hypothetical prote
12	94.5	10.1	883	2 A84210	hypothetical prote
13	94	10.1	1090	2 T00533	probable DNA2-NAM7
14	93.5	10.0	2048	2 C84609	hypothetical prote
15	93	10.0	881	2 AD2580	two component sens
16	93	10.0	881	2 C97362	protein sensor pro
17	92.5	9.9	503	2 AH3535	periplasmic dipept
18	89	9.5	955	2 A47334	Lckin kinesin-rela
19	88	9.4	500	2 B81060	N utilisation subs
20	88	9.4	505	2 H81816	N utilisation subs
21	88	9.4	775	2 T45238	probable transfera
22	87	9.3	469	2 B87094	probable molecular
23	86.5	9.3	157	2 F84349	hypothetical prote
24	85	9.1	2703	2 H81193	hemagglutinin/hemo
25	84.5	9.1	266	2 C84365	conserved hypothet
26	84	9.0	491	2 F96022	cellulose synthase
27	83.5	8.9	712	2 T02552	gas-vesicle protei
28	82.5	8.8	381	2 S28115	readthrough protei
29	82	8.8	331	2 S01964	

ALIGNMENTS

RESULT 1

E84304

Htr10 transducer [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: E84304

R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S

; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo

Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li

A>Title: Genome sequence of Halobacterium species NRC-1.

A:Reference number: A84160; MUID:20504483; PMID:11016950

A:Accession: E84304

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-489 <STO>

A:Cross-references: GB:AE004437; NID:g10580997; PIDN:AAG19801.1; GSPDB:GN00138

C:Genetics:

A:Gene: htr10

C:Superfamily: Halobacterium salinarum transducer protein htr1

Query Match 100.0%; Score 933; DB 2; Length 489;
Best Local Similarity 100.0%; Pred. No. 1.9e-70;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT	60
Db	1	MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT	60
QY	61	ADALVTDFYDHLSEYRTQDLFANSTKTVEQLKQAEYLLGLGRGEYDTEYAQRARIG	120
Db	61	ADALVTDFYDHLSEYRTQDLFANSTKTVEQLKQAEYLLGLGRGEYDTEYAQRARIG	120
QY	121	KIHVLGLGPDVYLGAVTRYTTGLLDALDDVADRGEEAAAADVELVARFLPMLKLITF	180
Db	121	KIHDLVLGPDVYLGAVTRYTTGLLDALDDVADRGEEAAAADVELVARFLPMLKLITF	180
QY	181	DQOI 184	
Db	181	DQOI 184	

RESULT 2

T44978

transducer protein hemAT [validated] - Halobacterium salinarum

N:Alternate names: methyl-accepting taxis protein htB; transducer protein htB; transducer

C:Species: Halobacterium salinarum

C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 15-Sep-2000

C:Accession: T44978

R:Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.

Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996

A;Title: Signal transduction in the archaeon Halobacterium salinarum is processed through the transducer protein YnfV - Bacillus subtilis
A;Reference number: 222804; MUID:96209786; PMID:8643459
A;Accession: T44978
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-489 <ZHA>
A;Cross-references: EMBL:U75436; NID:g1654420; PIDN:AAB17981.1; PID:g1654421
A;Experimental source: strain Fix15
A;Note: The source is designated as Halobacterium salinarum
C;Genetics:
A;Gene: hemAT; htp15; htb
C;Function:
A;Description: involved in aerotactic signal transduction; involved in oxygen sensing; involved in signal transduction
C;Superfamily: Halobacterium salinarum transducer protein htrI
C;Keywords: heme; methylated amino acid; signal transduction

Query Match 100.0%; Score 933; DB 2; Length 489;
Best Local Similarity 100.0%; Pred. No. 1.9e-70; Mismatches 0; Indels 0; Gaps 0;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
Db |||||
QY 1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
Db |||||

QY 61 ADALVTDFVHLESYERTQDLFANSTKVQKETAQAEVLLGLGRGEYDTEVAARIG 120
Db |||||
QY 61 ADALVTDFVHLESYERTQDLFANSTKVQKETAQAEVLLGLGRGEYDTEVAARIG 120
Db |||||

QY 121 KHDVILGDPVYLGAYTRYTGLDALADDVVDGRGEAAAVDELVARFLPMLKLTFF 180
Db |||||
QY 121 KHDVILGDPVYLGAYTRYTGLDALADDVVDGRGEAAAVDELVARFLPMLKLTFF 180
Db |||||

QY 181 DQOI 184
Db |||||
QY 181 DQOI 184
Db |||||

RESULT 3
A83713
methyl-accepting chemotaxis protein BH0505 [imported] - Bacillus halodurans (strain C-12)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: A83713
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-439 <STO>
A;Cross-references: GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BA04224.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH0505

Query Match 15.9%; Score 148; DB 2; Length 439;
Best Local Similarity 25.2%; Pred. No. 8e-05; Mismatches 35; Indels 10; Gaps 3;
Matches 39; Conservative 35; Mismatches 71; Indels 10; Gaps 3;

QY 30 EARIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFVHLESYERTQDLFANSTKV 89
Db |||||
QY 30 EARIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFVHLESYERTQDLFANSTKV 89
Db |||||

QY 90 EQLKETQAEVLLGLGRGEYDTEVAARIGKTHDVLGDPVYLGAYTRYTGLDALA 149
Db |||||
QY 90 EQLKETQAEVLLGLGRGEYDTEVAARIGKTHDVLGDPVYLGAYTRYTGLDALA 149
Db |||||

QY 150 DVVADGRGEAAAVDELVARFLPMLKLTFFDQOI 184
Db |||||
QY 150 DVVADGRGEAAAVDELVARFLPMLKLTFFDQOI 184
Db |||||

QY 144 DSLI-QLLEQHLOSPSDIVLATRSLLKLNLEQOL 177
Db |||||

RESULT 4

C69832

methyl-accepting chemotaxis protein homolog ynfV - Bacillus subtilis

C;Species: Bacillus subtilis

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C;Accession: C69832

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero

A;Authors: Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero

A;Reference number: 222804; MUID:96209786; PMID:8643459

A;Accession: T44978

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-489 <ZHA>

A;Cross-references: EMBL:U75436; NID:g1654420; PIDN:AAB17981.1; PID:g1654421

A;Experimental source: strain Fix15

A;Note: The source is designated as Halobacterium salinarum

C;Genetics:

A;Gene: hemAT; htp15; htb

C;Function:

A;Description: involved in aerotactic signal transduction; involved in oxygen sensing; involved in signal transduction

C;Superfamily: Halobacterium salinarum transducer protein htrI

C;Keywords: heme; methylated amino acid; signal transduction

Query Match 100.0%; Score 933; DB 2; Length 489;
Best Local Similarity 100.0%; Pred. No. 1.9e-70; Mismatches 0; Indels 0; Gaps 0;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
Db |||||QY 1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
Db |||||QY 61 ADALVTDFVHLESYERTQDLFANSTKVQKETAQAEVLLGLGRGEYDTEVAARIG 120
Db |||||QY 61 ADALVTDFVHLESYERTQDLFANSTKVQKETAQAEVLLGLGRGEYDTEVAARIG 120
Db |||||QY 121 KHDVILGDPVYLGAYTRYTGLDALADDVVDGRGEAAAVDELVARFLPMLKLTFF 180
Db |||||QY 121 KHDVILGDPVYLGAYTRYTGLDALADDVVDGRGEAAAVDELVARFLPMLKLTFF 180
Db |||||QY 181 DQOI 184
Db |||||QY 181 DQOI 184
Db |||||

RESULT 3

A83713

methyl-accepting chemotaxis protein BH0505 [imported] - Bacillus halodurans (strain C-12)

C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C;Accession: A83713

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A;Reference number: A83650; MUID:20512582; PMID:11058132

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-439 <STO>

A;Cross-references: GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BA04224.1; GSPDB:GN00

A;Experimental source: strain C-125

C;Genetics:

A;Gene: BH0505

Query Match 15.9%; Score 148; DB 2; Length 439;
Best Local Similarity 25.2%; Pred. No. 8e-05; Mismatches 35; Indels 10; Gaps 3;
Matches 39; Conservative 35; Mismatches 71; Indels 10; Gaps 3;QY 30 EARIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFVHLESYERTQDLFANSTKV 89
Db |||||QY 30 EARIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFVHLESYERTQDLFANSTKV 89
Db |||||QY 90 EQLKETQAEVLLGLGRGEYDTEVAARIGKTHDVLGDPVYLGAYTRYTGLDALA 149
Db |||||QY 90 EQLKETQAEVLLGLGRGEYDTEVAARIGKTHDVLGDPVYLGAYTRYTGLDALA 149
Db |||||QY 150 DVVADGRGEAAAVDELVARFLPMLKLTFFDQOI 184
Db |||||QY 150 DVVADGRGEAAAVDELVARFLPMLKLTFFDQOI 184
Db |||||QY 144 DSLI-QLLEQHLOSPSDIVLATRSLLKLNLEQOL 177
Db |||||

RESULT 4

C69832

methyl-accepting chemotaxis protein homolog ynfV - Bacillus subtilis

C;Species: Bacillus subtilis

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C;Accession: C69832

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero

A;Authors: Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero

A;Reference number: 222804; MUID:96209786; PMID:8643459

A;Accession: T44978

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-489 <ZHA>

A;Cross-references: EMBL:U75436; NID:g1654420; PIDN:AAB17981.1; PID:g1654421

A;Experimental source: strain Fix15

A;Note: The source is designated as Halobacterium salinarum

C;Genetics:

A;Gene: hemAT; htp15; htb

C;Function:

A;Description: involved in aerotactic signal transduction; involved in oxygen sensing; involved in signal transduction

C;Superfamily: Halobacterium salinarum transducer protein htrI

C;Keywords: heme; methylated amino acid; signal transduction

Query Match 100.0%; Score 933; DB 2; Length 489;
Best Local Similarity 100.0%; Pred. No. 1.9e-70; Mismatches 0; Indels 0; Gaps 0;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
Db |||||QY 1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
Db |||||QY 61 ADALVTDFVHLESYERTQDLFANSTKVQKETAQAEVLLGLGRGEYDTEVAARIG 120
Db |||||QY 61 ADALVTDFVHLESYERTQDLFANSTKVQKETAQAEVLLGLGRGEYDTEVAARIG 120
Db |||||QY 121 KHDVILGDPVYLGAYTRYTGLDALADDVVDGRGEAAAVDELVARFLPMLKLTFF 180
Db |||||QY 121 KHDVILGDPVYLGAYTRYTGLDALADDVVDGRGEAAAVDELVARFLPMLKLTFF 180
Db |||||QY 181 DQOI 184
Db |||||QY 181 DQOI 184
Db |||||

RESULT 3

A83713

methyl-accepting chemotaxis protein BH0505 [imported] - Bacillus halodurans (strain C-12)

C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C;Accession: A83713

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A;Reference number: A83650; MUID:20512582; PMID:11058132

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-439 <STO>

A;Cross-references: GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BA04224.1; GSPDB:GN00

A;Experimental source: strain C-125

C;Genetics:

A;Gene: BH0505

Query Match 15.9%; Score 148; DB 2; Length 439;
Best Local Similarity 25.2%; Pred. No. 8e-05; Mismatches 35; Indels 10; Gaps 3;
Matches 39; Conservative 35; Mismatches 71; Indels 10; Gaps 3;QY 30 EARIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFVHLESYERTQDLFANSTKV 89
Db |||||QY 30 EARIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFVHLESYERTQDLFANSTKV 89
Db |||||QY 90 EQLKETQAEVLLGLGRGEYDTEVAARIGKTHDVLGDPVYLGAYTRYTGLDALA 149
Db |||||QY 90 EQLKETQAEVLLGLGRGEYDTEVAARIGKTHDVLGDPVYLGAYTRYTGLDALA 149
Db |||||QY 150 DVVADGRGEAAAVDELVARFLPMLKLTFFDQOI 184
Db |||||QY 150 DVVADGRGEAAAVDELVARFLPMLKLTFFDQOI 184
Db |||||QY 144 DSLI-QLLEQHLOSPSDIVLATRSLLKLNLEQOL 177
Db |||||

QY 33 IAWRLSFTGIDDDTAAALAEQPLFEATADALVTDYDHLSEYERTQDLFANSTKTVEQL 92
 Db 11 IGERTAFMGIDDKARSALRDLRPVIRAEIGKALDNEYGKVRATPFRKFFSDRRHMAAS 70
 QY 93 KETQAEYLLGLGRGEYDTEYAQRARIGKIHVDVLGPDVYLGAYTRYTGLLDALADV 152
 Db 71 SRQQAHWGV-IAEQPSDDYQAVRAIGQTHARIGLEPRWYIGGYAVVGDHLVRAVDSM 129
 QY 153 -----VADRGEEAAAADVELVARELPM 174
 Db 130 WPRGLLAKGSDRAGEAVAL--MKAIFLDM 158
 RESULT 6
 A:Map position: circular chromosome
 A:Query Match 12.6%; Score 117.5; DB 2; Length 499;
 Best Local Similarity 26.5%; Pred. No. 0.034;
 Matches 35; Conservative 23; Mismatches 67; Indels 7; Gaps 3;
 methyl-accepting chemotaxis protein mcpV (AF312877) [imported] - Agrobacterium tumefaciens
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
 C:Accession: A97485
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: A97485
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-499 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK86834.1; PID:g15156046; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C1888
 A:Map position: circular chromosome
 Query Match 12.6%; Score 117.5; DB 2; Length 499;
 Best Local Similarity 26.5%; Pred. No. 0.034;
 Matches 35; Conservative 23; Mismatches 67; Indels 7; Gaps 3;
 QY 19 GHALADRIGLDEAEIAWRLSFTGIDDDTAAALAEQPLFEATADALVTDYDHLSEYERT 78
 Db 3 GOAKTDR-QLDE-----RLNFLGLGHERQNLSDMKGVITGSLDASLDRFYTKVRAVPET 56
 QY 79 QDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAQRARIGKIHVDVLGPDVYLGAYT 138
 Db 57 AKFF-SSEAHIIHAKSMQLKHSRIASGTFFNEDYTNVTAIGRTHARLGLEPRWYIGGYA 115
 QY 139 RYVTGLLDALAD 150
 Db 116 LMLDGIVKAVIE 127
 RESULT 7
 A:Map position: circular chromosome
 A:Query Match 12.6%; Score 117.5; DB 2; Length 499;
 Best Local Similarity 26.5%; Pred. No. 0.034;
 Matches 35; Conservative 23; Mismatches 67; Indels 7; Gaps 3;
 methyl-accepting chemotaxis protein mcpV [imported] - Agrobacterium tumefaciens (strain
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C:Accession: AB2703
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AB2703
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-499 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AAAL42040.1; PID:g17739417; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: mcpV

A:Map position: circular chromosome

Query Match 12.6%; Score 117.5; DB 2; Length 499;
 Best Local Similarity 26.5%; Pred. No. 0.034;
 Matches 35; Conservative 23; Mismatches 67; Indels 7; Gaps 3;
 QY 19 GHALADRIGLDEAEIAWRLSFTGIDDDTAAALAEQPLFEATADALVTDYDHLSEYERT 78
 Db 3 GOAKTDR-QLDE-----RLNFLGLGHERQNLSDMKGVITGSLDASLDRFYTKVRAVPET 56
 QY 79 QDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAQRARIGKIHVDVLGPDVYLGAYT 138
 Db 57 AKFF-SSEAHIIHAKSMQLKHSRIASGTFFNEDYTNVTAIGRTHARLGLEPRWYIGGYA 115
 QY 139 RYVTGLLDALAD 150
 Db 116 LMLDGIVKAVIE 127

RESULT 8

D87536
 methyl-accepting chemotaxis protein McpM [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: D87536
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapero, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: D87536
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-555 <STO>
 A:Cross-references: GB:AE005673; NID:g13423838; PIDN:AAK24288.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC2317
 Query Match 11.3%; Score 105.5; DB 2; Length 555;
 Best Local Similarity 24.7%; Pred. No. 0.39;
 Matches 40; Conservative 28; Mismatches 77; Indels 17; Gaps 5;
 QY 31 AEIAWRLSFTGIDDDTAAALAEQPLFEATADALVTDYDHLSEYERTQDLFANSTKTVE 90
 Db 46 AKLDORMAFMFRDERSRAHLFAIKPEVIDAEIAGAALGFYSQVFLFPDTRVKFRDDGH-MA 104
 QY 91 QLKETQAEYLLGLGRGEYDTEYAQRARIGKIHVDVLGPDVYLGAYTRYTGLLDALAD 150
 Db 105 GAERAQAQAHWRRIAEAGYGSYVDVERIGRSHADADIAPOWYIGGYAVVVEEVMRAL-- 162
 QY 151 DVVADRG-----EAAAAYDELV-ARFLPM-LKLITF 180
 Db 163 --VAKRAKGLFNSAKSDAELADGLSALIKAAFLMDLSVSTY 202
 RESULT 9
 19G-binding protein - Streptococcus sp. (group G)
 C:Species: Streptococcus sp.
 C:Date: 17-Sep-1987 #sequence_revision 17-Sep-1987 #text_change 17-Mar-2000
 C:Accession: A24496; A39041
 R:Fainestock, S.R.; Alexander, P.; Nagle, J.; Filpula, D.
 J. Bacteriol. 167, 870-880, 1986
 A:Title: Gene for an immunoglobulin-binding protein from a group G Streptococcus.
 A:Reference number: A24496; MUID:86304178; PMID:3745123
 A:Accession: A24496
 A:Molecule type: DNA
 A:Residues: 1-448 <FAH>
 A:Cross-references: GB:M13825; NID:g153822; PIDN:AAA03664.1; PID:g153823
 R:Sjoerding, U.; Bjoerck, L.; Kastern, W.
 J. Biol. Chem. 266, 399-405, 1991
 A:Title: Streptococcal protein G. Gene structure and protein binding properties.

A:Reference number: A39041; MUID:91093154; PMID:1985908
A:Accession: A39041
A>Status: preliminary
A:Molecule type: protein
A:Residues: 34-42, 'N', 45-48, 62-76, 186-200 <SUO>
C:Genetics:
A:Gene: spg
C:Superfamily: M5 protein
C:Keywords: transmembrane protein

Query Match 10.6%; Score 99; DB 2; Length 448;
Best Local Similarity 25.3%; Pred. No. 1;
Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;
QY 10 TADVRNGIDGHADRIGLDEAEIAWRLSFTGIDD-----DTMAALAAE----- 53
Db 41 TPIIRNGGE--LTNLLGNSETTLARNEESATADLTAAAVADTVAAAAAENAGAAWEEA 97
QY 54 -----OPLFEATADAL-----VTFDYHLESYERTQDLFANSTKTVEQLKETQAEYLLG 102
Db 98 AAADALAKAKADALKFNKYGVSYYKNL-----INNKTVEGKDLQAOVW-- 145
QY 103 LRGEYDTEYAAQRAIRIGKHIDVGLGPDVYLGAIVRYVYTGLLDALADDVADRGEEAAA 162
Db 146 -----ESAKKARISEATD-----GLSDFLKSQTPA---EDTVK 175
QY 163 AVDELVARFL 172
Db 176 SIELAEAKVL 185

RESULT 10
S00128
protein G precursor - Streptococcus sp. (Streptococcus G148)
N:Alternate names: albumin-binding protein; cell wall-bound protein
C:Species: Streptococcus sp.
A:Variety: Streptococcus G148
C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 19-May-2000
C:Accession: S00128; A27604; A26314
R:Olsson, A.; Eliasson, M.; Guss, B.; Nilsson, B.; Hellman, U.; Lindberg, M.; Uhlen, M.
Eur. J. Biochem. 168, 319-324, 1987
A:Title: Structure and evolution of the repetitive gene encoding streptococcal protein G
A:Reference number: S00128; MUID:88029445; PMID:3665928
A:Accession: S00128
A:Molecule type: DNA
A:Residues: 1-593 <OLS>
A:CROSS-references: EMBL:X06173; NID:G47084; PIDN:CAA29540.1; PID:G47085
A:Note: the source is designated as Streptococcus G148
A:Note: part of this sequence, including the amino end of the mature protein, was confir
J. Sjoerding, U.; Falkenberg, C.; Nielsen, E.; Akerstroem, B.; Bjoerck, L.
J. Immunol. 140, 1595-1599, 1988
A:Title: Isolation and characterization of a 14-kDa albumin-binding fragment of streptoc
A:Reference number: A27604; MUID:88154455; PMID:2831269
A:Accession: A27604
A:Molecule type: protein
A:Residues: 62-101 <SUO>
R:Guss, B.; Eliasson, M.; Olsson, A.; Uhlen, M.; Frej, A.K.; Jorvall, H.; Flock, J.I.;
EMBO J. 5, 1567-1575, 1986
A:Title: Structure of the IgG-binding regions of streptococcal protein G.
A:Reference number: A26314; MUID:86300657; PMID:3017704
A:Accession: A26314
A:Molecule type: DNA
A:Residues: 114-593 <GUS>
A:CROSS-references: GB:X04015; NID:G47071; PIDN:CAA27638.1; PID:G47072
C:Function:
A:Description: it is part of the cell wall structure of group G streptococci and is cova
C:Superfamily: M5 protein
C:Keywords: duplication; membrane protein
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-593/Product: protein G #status experimental <MAT>
F:34-116/Domain: alanine-rich <ALA>
F:117-290/Domain: AB duplication <DUPL>
F:117-140/Region: A repeat

F:141-191/Region: B
F:192-215/Region: A repeat
F:216-266/Region: B
F:267-290/Region: A repeat
F:303-497/Domain: IgG binding <IGB>
F:303-357/Region: C repeat
F:358-372/Region: D
F:373-427/Region: C repeat
F:428-442/Region: D
F:443-497/Region: C repeat
F:498-567/Domain: proline-rich <PRO>
F:568-593/Domain: carboxyl-terminal <CTD>

Query Match 10.5%; Score 98; DB 2; Length 593;
Best Local Similarity 25.3%; Pred. No. 1.8;
Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;
QY 10 TADVRNGIDGHADRIGLDEAEIAWRLSFTGIDD-----DTMAALAAE----- 53
Db 41 TPIIRNGGE--LTNLLGNSETTLARNEESATADLTAAAVADTVAAAAAENAGAAWEEA 97
QY 54 -----OPLFEATADAL-----VTFDYHLESYERTQDLFANSTKTVEQLKETQAEYLLG 102
Db 98 AAADALAKAKADALKFNKYGVSYYKNL-----INNKTVEGKDLQAOVW-- 145
QY 103 LRGEYDTEYAAQRAIRIGKHIDVGLGPDVYLGAIVRYVYTGLLDALADDVADRGEEAAA 162
Db 146 -----ESAKKARISEATD-----GLSDFLKSQTPA---EDTVK 175
QY 163 AVDELVARFL 172
Db 176 SIELAEAKVL 185

RESULT 11
T35174
hypothetical protein SC5A7.25c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 04-Mar-2000
C:Accession: T35174
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1998
A:Reference number: Z21570
A:Accession: T35174
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-218 <SEE>
A:CROSS-references: EMBL:AL031107; PIDN:CAA19954.1; GSPDB:GN00070; SCOEDB:SC5A7.25c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC5A7.25c
C:Superfamily: Streptomyces coelicolor hypothetical protein SC5A7.25c

Query Match 10.4%; Score 97; DB 2; Length 218;
Best Local Similarity 29.6%; Pred. No. 0.62;
Matches 53; Conservative 21; Mismatches 77; Indels 28; Gaps 10;
QY 7 TLVADVVRNGIDGHADRIGLDEAEI--AWRLS-PTGIDDDTMAALAAEQPLFEATADA 63
Db 15 TLAADVVRVG-----ALADRLDVPHAEVFDVGRLSAASGVPSVVGALLGRP-----TGGA 66
QY 64 LV-TDFYHLESYERTQDLFANSTKTVEQ-----LKETQAEYLLGLGRGEYDTEYAAQ 115
Db 67 DVQTRFVRQDLRLRRTR-LKPNGRKKTQEIADGAGMSRQQAGALIN-GURRPTMEHCDA 124
QY 116 RARIGKIDH--VLGPGDVYLGAIVRYVYTGLLDALADDVADRGEEAAAANDELVARFL 172
Db 125 IQRFFRVHAGFLTAEDPEALAGALQRTQELQQL-----ADREQAAPADPLRL 178

RESULT 12
A84210
hypothetical protein Vng0514c [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: AB4210
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leitchauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freltas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1
A:Reference number: AB4160; MUID:20504483; PMID:11016950
A:Accession: AB4210
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-883 <STO>
A:Cross-references: GB:AE004437; NID:g10580117; PIDN:AAG19045.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG0514C

Query Match 10.1%; Score 94.5; DB 2; Length 883;
Best Local Similarity 28.3%; Pred. No. 5.8;
Matches 51; Conservative 23; Mismatches 63; Indels 43; Gaps 10;

QY 12 DVNRGIDGH--ALADRGIDGAEIAWRLSFTGIDDDTMAALAAEQPLFEA-----TAD 62
DB 180 DVKSVEGQDLRLADQIA--DKENADPHDLASHNTALAEVTAIDIEHFEAEREQARQTRD 237
QY 63 ALVTDYFHLESYERTQDLFANSTKTVEQLKETOAEYLLGLGRGYDTEYAAQRA----- 117
DB 238 ----DAADVLEYEBSRTALADVETIADVRAVE-----ABRETTLADRVSDHRE 286
QY 118 RIGKHIV-----LGL-GPDVYLGATRYTYTGLLDALAD--DVVADRGEEAAAADVEL 167
DB 287 RASLDLDEAAALADLGLDDPAEDASAER-----DAVADQREAVREVPVAVSRL 340

RESULT 13
T00533
probable DNA2-NAM7 helicase family protein [imported] - Arabidopsis thaliana
N:Alternate names: SEN1 protein homolog T20K24.14
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Feb-2001
C:Accession: T00533; G84572
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, July 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20K24 genomic sequence.
A:Reference number: Z14167
A:Accession: T00533
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1090 <ROU>
A:Cross-references: EMBL:AC002392; NID:g3176701; PID:g3176714
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:10617197
A:Accession: G84572
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1090 <STO>
A:Cross-references: GB:AE002093; NID:g3176714; PIDN:AAD12029.1; GSPDB:GN00139
C:Genetics:
A:Gene: T20K24.14; At2g919120
A:Map position: 2
A:Introns: 250/1; 310/1; 342/1; 383/2; 427/3; 1016/3

Query Match 10.1%; Score 94; DB 2; Length 1090;
Best Local Similarity 26.9%; Pred. No. 8.3;
Matches 39; Conservative 17; Mismatches 55; Indels 34; Gaps 5;

QY 55 PLFE----ATADALVTDFYHLESYERTQDLFA-----NSTKTVEQLKETOAEYLLGL 103

C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: AB4210
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leitchauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freltas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1
A:Reference number: AB4160; MUID:20504483; PMID:11016950
A:Accession: AB4210
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-883 <STO>
A:Cross-references: GB:AE004437; NID:g10580117; PIDN:AAG19045.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG0514C

Query Match 10.1%; Score 93.5; DB 2; Length 2048;
Best Local Similarity 26.3%; Pred. No. 20;
Matches 55; Conservative 25; Mismatches 78; Indels 51; Gaps 11;

QY 3 NDNDTLVTADVNR-NGID--GHALADRIGL-----DEAEIARLSTFTGIDDDTMAALAAEQP 55
DB 1302 NPSRALAVADVEMNAVDVLCRLSSNTYTMELKGDAELCYVLFANTRIRSTVAAACVPEP 1361
QY 56 LFEATADALVTDFYHLESYERTQDLFANSTKTVEQLKETOAEY-----LLGLGRGE-YD 109
DB 1362 L-----VSLVTVEFSPAQSHVVRAIDKLVDV-----EQALVAAHGAHVPLVGLLYGKNYV 1413
QY 110 TEYAAQRA--RIGK-----IHDVGLGPDVYLGATRYTYTGLLDALAD 150
DB 1414 LHEAISRALVKLGKDRPACKLEMYKAGVIDCVLDILHEAPDFLCAAFSE-----LLRLITN 1469
QY 151 DVVADRGEEAAAADVELVARFLPMLKLLT 179
DB 1470 NATIAGQSAKAVVE-----PLFHLLT 1491

RESULT 15
AD2580
two component sensor kinase Atu0027 [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AD2580
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClella
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD2580
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-881 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL41058.1; PID:g17738345; GSPDB:GN00186

A; Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: Atu0027

A;Map position: circular chromosome

	Query Match	10.0%;	Score 93;	DB 2;	Length 881;
	Best Local Similarity	28.7%;	Pred. No. 7.7;		
	Matches	49;	Conservative	23;	Mismatches 57; Indels 42; Gaps 10;
QY	37	LSFTGTDGDDTMAALAAEQPLFEAT----	ADALVTDFYDHLSEYVERTQDL-----	PANST	86
DB	621	LTFVNMTD----	SVRAERALKKNDALLKADELKNDVFQHV--SYELSPNTNIIGFTDLL		675
QY	87	KT--VEQLKETQAEYLLGLGRGEYDTETAAQRAARIGKIHVDVLGPGDPVYLQ--AYTRYTGT			143
DB	676	KTPGIGQLTERQAEYL-----	DHISTSSVLLTTIVNDILDLA--TVDAGIMQLNYSDN		726
QY	144	LILDALADV---	VADRGEEAAAAVDELVARFLP-----	MUKLIT	179
DB	727	DLNELLDGDSVOIADRLQESGISELIVAPAHGLGSVADHORLQKILFKLIT			777

Search completed: August 10, 2004, 15:29:28
Job time : 9.01937 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:19:00 ; Search time 5.79177 Seconds
(without alignments)
1654.229 Million cell updates/sec

Title: US-09-455-978B-77

Perfect score: 933

Sequence: 1 MSNDNDTLVTADVRNGIDGH.....DELVARFLPMLKLLTFDQOI 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	933	100.0	489	1	HMAT_HALN1	Q9hpr6	halobacteri
2	933	100.0	489	1	HMAT_HALSA	P71413	halobacteri
3	142	15.2	432	1	HMAT_BACSU	O07621	bacillus su
4	99	10.6	448	1	SPG1_STRSG	P06654	streptococ
5	98	10.5	593	1	SPG2_STRSG	P19909	streptococ
6	90.5	9.7	305	1	LPXC_RALSO	Q9xhl6	raistonia s
7	89	9.5	955	1	KINL_LEICH	P46865	leishmania
8	88	9.4	775	1	PLSB_MYCLE	Q9X7B0	mycobacteri
9	87	9.3	469	1	TIG_MYCLE	Q9Cby2	mycobacteri
10	85.5	9.2	957	1	SECA_MYCSM	P71533	mycobacteri
11	82.5	8.8	381	1	GVPC_HALME	O02228	halobacteri
12	82	8.8	331	1	VAL_BPSP	P09677	bacterioph
13	82	8.8	372	1	PROB_PSEAE	Q9hvl9	pseudomonas
14	81	8.7	1001	1	IF2_SYNY3	P72689	synecocyst
15	80.5	8.6	764	1	HTR2_HALSA	P71410	halobacteri
16	80.5	8.6	765	1	HTR2_HALSA	P71410	halobacteri
17	80.5	8.6	765	1	HTR2_HALSA	P71410	halobacteri
18	80.5	8.6	765	1	HTR2_HALSA	P71410	halobacteri
19	80	8.6	867	1	HTR2_HALSA	Q9n2i7	felis silve
20	80	8.6	461	1	K6PF_THETI	Q9rfj8	fusobacteri
21	79.5	8.5	934	1	K6PF_PSESM	Q9hnl1	thermococcu
22	79.5	8.5	334	1	YU03_RHIME	Q98715	pseudomonas
23	79.5	8.5	1415	1	HTR2_HALN1	Q92v44	rhizobium m
24	78.5	8.4	374	1	PROB_RALSO	Q9h8p1	haemophilus
25	78.5	8.4	382	1	GVCL_HALN1	Q9xvll	raistonia s
26	78.5	8.4	454	1	K6PF_PYRPU	P24574	halobacteri
27	78.5	8.4	584	1	Y742_STRCO	Q9v2z7	pyrococcus
28	78.5	8.4	766	1	DPP4_HUMAN	Q9znbo	streptomyce
29	78.5	8.4	949	1	SEAL_MYCTU	P27487	homo sapien
30	78	8.4	442	1	HRB6_XANCY	O05885	mycobacteri
31	78	8.4	450	1	K6PF_PYRHO	P80153	xanthomonas
32	77.5	8.3	459	1	K6PF_THELI	O59355	pyrococcus
33	77	8.3	323	1	T2B1_BACST	P70985	bacillus st

34	77	8.3	372	1	PROB_PSESM	Q889f0	pseudomonas
35	77	8.3	571	1	V096_FOWPV	Q9j5c4	fowlpox vir
36	77	8.3	591	1	IF2F_METMA	Q8pu78	methanosarc
37	76.5	8.2	451	1	K6PF_PYRAB	Q9vla6	pyrococcus
38	76.5	8.2	483	1	SYC_SYNY3	P74330	synecocyst
39	76.5	8.2	495	1	ENGR_YERPE	Q8zct9	yersinia pe
40	76.5	8.2	753	1	CAT2_NEUCR	Q8xi82	neurospora
41	76.5	8.2	940	1	SECA_STRGR	P95759	streptomyce
42	76.5	8.2	1683	1	YJD9_YEAST	P47054	saccharomyce
43	76	8.1	171	1	RS13_HALN1	Q9v2w4	halobacteri
44	76	8.1	239	1	VG88_BPML5	Q05305	mycobacteri
45	76	8.1	654	1	SYT_LACPL	Q88wu9	lactobacilli

ALIGNMENTS

RESULT 1

HMAT_HALN1
ID HMAT_HALN1 STANDARD; PRT; 489 AA.
AC Q9HPR6;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heme-based aerotactic transducer HEMAT.
GN HEMAT OR HTR10 OR VNG1505G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddock D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.,
RA "Genome sequence of Halobacterium species NRC-1.";
Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
CC -!- FUNCTION: Heme-containing signal transducer responsible for
aerotaxis, the migratory response toward or away from oxygen (By
similarity).
CC -!- SUBUNIT: Homotetramer (Probable).
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
or send an email to license@isb-sib.ch).

EMBL; AE005064; AAC19801.1; -.
DR PIR; E84304; E84304.
DR HSSP; P02942; 1QV7.
DR InterPro; IPR004089; Chtaxis_trans.
DR InterPro; IPR004090; Met_chemotaxis.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCH.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
KW Transducer; Heme; Complete proteome.
FT DOMAIN 218 454 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 489 AA; 52852 MW; 8C0AEF179667791E CRC64;

Query Match 100.0%; Score 933; DB 1; Length 489;
Best Local Similarity 100.0%; Pred. No. 4.1e-68;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

KW Transducer; Heme.
PT DOMAIN 218 454 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 489 AA; 52816 MW; 8457263FCF616BFF CRC64;

Query Match 100.0%; Score 933; DB 1; Length 489;
Best local Similarity 100.0%; Pred. No. 4.1e-68;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNDNDTLVTADVNRGIDGHALADRIGLDEAIEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
Db |||||
1 MSNDNDTLVTADVNRGIDGHALADRIGLDEAIEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60

QY 61 ADALVTDFYDHLSEYERTQDLFANSTKVEQLKETOAEYLLGLGRGEYDTYAAQARIG 120
Db |||||
61 ADALVTDFYDHLSEYERTQDLFANSTKVEQLKETOAEYLLGLGRGEYDTYAAQARIG 120

QY 121 KIHVDVLGPDVYLGAITYRYTGLLDALADVDVADRGEEAAAANDELVARFLPMLKLTFF 180
Db |||||
121 KIHVDVLGPDVYLGAITYRYTGLLDALADVDVADRGEEAAAANDELVARFLPMLKLTFF 180

QY 181 DQOI 184
Db |||||
181 DQOI 184

RESULT 3
HMAT_BACSU
ID HMAT_BACSU STANDARD; PRT; 432 AA.
AC C07621;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Heme-based aerotactic transducer hemAT.
OS HEMAT OR BSU10380.
GN Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]_TaxID=1423;
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98240224; PubMed=9579061;
RA Noback M.A., Holsappel S., Kiewiet R., Terpstra P., Wambutt R.,
  Wedler H., Venema G., Bron S.;
RT "The 172 kb prkA-addAB region from 83 degrees to 97 degrees of the
  Bacillus subtilis chromosome contains several dysfunctional genes,
  the glyB marker, many genes encoding transporter proteins, and the
  ubiquitous hit gene.";
RL Microbiology 144:859-875(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
  Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
  Borriss R., Bourret L., Brans A., Braun M., Brignell S.C., Bron S.,
  Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
  Choi S.K., Codani J.J., Connerston I.F., Cummings N.J., Daniel R.A.,
  Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
  Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
  Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Gaileron N.,
  Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grand G.,
  Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
  Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
  Joris B., Karamata D., Kasahara Y., Koningstein G., Krogh S., Kumano M.,
  Kobayashi Y., Koetter P., Koningstein G., Lauber J., Lazarevic V.,
  Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
  Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
  Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
  Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
  Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
  Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
  Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
  Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
```

```

QY 1 MSNDNDTLVTADVNRGIDGHALADRIGLDEAIEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
Db |||||
1 MSNDNDTLVTADVNRGIDGHALADRIGLDEAIEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60

QY 61 ADALVTDFYDHLSEYERTQDLFANSTKVEQLKETOAEYLLGLGRGEYDTYAAQARIG 120
Db |||||
61 ADALVTDFYDHLSEYERTQDLFANSTKVEQLKETOAEYLLGLGRGEYDTYAAQARIG 120

QY 121 KIHVDVLGPDVYLGAITYRYTGLLDALADVDVADRGEEAAAANDELVARFLPMLKLTFF 180
Db |||||
121 KIHVDVLGPDVYLGAITYRYTGLLDALADVDVADRGEEAAAANDELVARFLPMLKLTFF 180

QY 181 DQOI 184
Db |||||
181 DQOI 184

RESULT 2
HMAT_HALSA
ID HMAT_HALSA STANDARD; PRT; 489 AA.
AC P71413;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heme-based aerotactic transducer hemAT (transducer Htb protein).
GN HEMAT OR HTB.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Flx15;
RX MEDLINE=96209786; PubMed=8643458;
RA Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
RT "Signal transduction in the archaeon Halobacterium salinarum is
  processed through three subfamilies of 13 soluble and membrane-bound
  transducer proteins.";
RL Nature 403:540-544(2000).
RN [2]
RP FUNCTION
RX MEDLINE=20140131; PubMed=10676961;
RA Hou S., Larsen R.W., Boudko D., Riley C.W., Karatan E., Zimmer M.,
  Ordal G.W., Alam M.;
RT "Myoglobin-like aerotaxis transducers in Archaea and Bacteria.";
RL Nature 403:540-544(2000).
CC -!- FUNCTION: Heme-containing signal transducer responsible for
  aerotaxis, the migratory response toward or away from oxygen.
CC -!- SUBUNIT: Homotetramer (Probable).
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U75436; AAB17881.1; --
CC PIR; T44978; T44978.
CC HSP; P02942; 1Q07.
CC GO; GO:0020037; F:heme binding; IDA.
CC GO; GO:0004871; F:signal transducer activity; IDA.
CC GO; GO:0009454; P:aerotaxis; IDA.
CC GO; GO:0007165; P:signal transduction; IDA.
CC InterPro; IPR004089; C:htaxis_trans.
CC InterPro; IPR004090; M:chemotaxis.
CC Pfam; PF00015; M:CPsignal; 1.
CC PRINTS; PR00260; CHEMTRNSDUCR.
CC SMART; SM00283; MA; 1.
CC PROSITE; PS0111; CHEMOTAXIS_TRANSDUC_2; 1.
```

RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitenegger T.,
RA Winters P., Wipat K., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RL Nature 390:249-256 (1997).
RN [3]
RN FUNCTION
RX MEDLINE=20140131; PubMed=10676961;
RA Hou S., Larsen R.W., Boudko D., Riley C.W., Karatan E., Zimmer M.,
RA Ordal G.W., Alam M.;
RT "Myoglobin-like aerotaxis transducers in *Archaea* and *Bacteria*.";
RL Nature 403:540-544 (2000).
RN [4]
RN RESONANCE RAMAN SPECTROSCOPY, AND SUBUNIT.
RX MEDLINE=21950695; PubMed=11821422;
RA Aono S., Kato T., Matsuki M., Nakajima H., Ohta T., Uchida T.,
RA Kitagawa T.;
RT "Resonance Raman and ligand binding studies of the oxygen-sensing
RT signal transducer protein HemaT from *Bacillus subtilis*.";
RL J. Biol. Chem. 277:13528-13538 (2002).
CC -!- FUNCTION: Heme-containing signal transducer responsible for
CC aerotaxis, the migratory response toward or away from oxygen.
CC -!- SUBUNIT: Homotetramer (Probable).
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Y14084; CAA74545.1; -;
DR EMBL; Z99109; CAB12878.1; -;
DR PIR; C69832; C69832.
DR Subtilist: BG13066; hemaT.
DR GO; GO:0020037; F:heme binding; IDA.
DR GO; GO:0004871; F:signal transducer activity; IDA.
DR GO; GO:0009454; P:aerotaxis; IDA.
DR GO; GO:0007165; P:signal transduction; IDA.
DR InterPro: IPR004089; Cmtaxis_transd.
DR Pfam: PF00015; MCPsignal; 1.
DR SMART: SM00283; MA; 1.
DR PROSITE; PS0111; CHEMOTAXIS_TRANSDUC_2; 1.
KW Transducer; Heme; Complete proteome.
FT DOMAIN 184 420 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 432 AA; 48767 MW; 87D0655974070B64 CRC64;
Query Match 15.2%; Score 142; DB 1; Length 432;
Best Local Similarity 25.3%; Pred. No. 0.00027;
Matches 39; Conservative 29; Mismatches 76; Indels 10; Gaps 2;
Qy 31 AEIAWRLSFTGIDDDTMAALAAEOPLEATDALVTDFYDHLSEYRTQDLPANSTKTVE 90
Db 32 ADVKQLMWELGDAELVLEQLQLEINIVNDAFYKNLHSHSLMDII-NDHSSVD 90
Qy 91 QLKETOAYLLGLGRGEYDTFYAQAQARIGKHIDVGLGPDVYLGAIVRYTYGLLDALAD 150
Db 91 RLKQTLKRHIQEMPAVGVIDDFIEKRNRIASIHRLIGLLPKWYMGAFQELLMSIDYI- 148
Qy 151 DVVADGEEAAAVDELVARFLPMLKLLTFDQOI 184
Db 149 -----EASITNQOELLKAIKATTKIINLEQOL 175
RESULT 4
SPGL_STRSG

ID AC SPGL_STRSG STANDARD; PRT; 448 AA.
DT P06654;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Immunoglobulin G binding protein G precursor (IGG binding protein G).
GN SPG.
OS Streptococcus sp. (Lancefield group G).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1320;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=86304178; PubMed=3745123;
RA Rahnestock S.R., Alexander P., Nagle J., Filpula D.;
RT "Gene for an immunoglobulin-binding protein from a group G
RT streptococcus.";
RL J. Bacteriol. 167:870-880 (1986).
RN [2]
RN X-RAY CRYSTALLOGRAPHY (2.07 ANGSTROMS) OF 228-282.
RX MEDLINE=94213848; PubMed=8161530;
RA Gallagher T., Alexander P., Bryan P., Gilliland G.L.;
RT "Two crystal structures of the B1 immunoglobulin-binding domain of
RT streptococcal protein G and comparison with NMR.";
RL Biochemistry 33:4721-4729 (1994).
RN [3]
RN X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 293-351.
RX MEDLINE=95055731; PubMed=7966308;
RA Derrick J.P., Wigley D.B.;
RT "The third IGG-binding domain from streptococcal protein G. An
RT analysis by X-ray crystallography of the structure alone and in a
RL complex with Fab.";
RN [4]
RN X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 297-352.
RX MEDLINE=95308043; PubMed=7788293;
RA Sauer-Briksson A.E., Kleywegt G.J., Uhlen M., Jones T.A.;
RT "Crystal structure of the C2 fragment of streptococcal protein G in
RL complex with the Fc domain of human IgG.";
RN [5]
RN X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 228-282.
RA Butterworth S., Lamzin V.S., Wigley D.B., Derrick J.P., Wilson K.S.;
RT Submitted (APR-1997) to the PDB data bank.
RN [6]
RN STRUCTURE BY NMR OF 298-351.
RX MEDLINE=91335209; PubMed=1871600;
RA Gronenborn A.M., Filpula D.R., Essig N.Z., Achari A., Whitlow M.,
RA Wingfield P.T., Clore G.M.;
RT "A novel, highly stable fold of the immunoglobulin binding domain of
RT streptococcal protein G.";
RL Science 253:657-661 (1991).
CC -!- FUNCTION: BINDS TO THE CONSTANT FC REGION OF IGG WITH HIGH
CC AFFINITY.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M13825; AAA03664.1; -;
DR PIR; A24496; A24496.
DR PDB; 1EM7; 08-MAY-02.
DR PDB; 1FCC; 20-JUL-95.
DR PDB; 1GBI; 15-APR-93.
DR PDB; 1IGG; 03-JUN-95.
DR PDB; 1IGD; 01-NOV-94.
DR PDB; 1MPE; 30-OCT-02.

DR PDB; 1MVK; 30-OCT-02.
DR PDB; 1PGA; 30-APR-94.
DR PDB; 1PGB; 30-APR-94.
DR PDB; 1PGX; 15-JUL-92.
DR PDB; 2GBL; 15-APR-93.
DR PDB; 2IGD; 29-JUL-98.
DR PDB; 2IGH; 31-JAN-94.
DR InterPro; IPR002988; GA.
DR InterPro; IPR005877; Gpos_Y5IRK.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000724; IgG_bind_B.
DR Pfam; PF01468; GA; 2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF01378; IgG_binding_B; 2.
DR Pfam; PF04650; Y5IRK_signal; 1.
DR PRINTS; PR00015; GPOSANCHOR.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs; TIGR01168; Y5IRK_signal; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
KW IgG-binding protein; Cell wall; Peptidoglycan-anchor; Repeat; Signal;
KW 3D-structure.
FT SIGNAL 1 33
FT CHAIN 34 417 IMMUNOGLOBULIN G BINDING PROTEIN G.
FT PROPEP 418 448 REMOVED BY SORTASE (POTENTIAL).
FT DOMAIN 69 111 ALA-RICH.
FT DOMAIN 104 215 2 X 37 AA REPEATS.
FT REPEAT 104 140 1-1.
FT REPEAT 179 215 1-2.
FT DOMAIN 228 352 2 X 55 AA REPEATS.
FT REPEAT 228 282 2-1.
FT REPEAT 298 352 2-2.
FT DOMAIN 386 410 5 X 5 AA REPEATS OF [DE]-D-A-K-K.
FT SITE 414 418 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 417 417 AMIDE-LINKED TO CELL WALL (POTENTIAL).
FT STRAND 228 234
FT STRAND 239 245
FT HELIX 249 262
FT TURN 263 264
FT STRAND 268 272
FT TURN 273 276
FT STRAND 277 281
FT STRAND 297 304
FT STRAND 309 316
FT HELIX 319 332
FT TURN 333 334
FT STRAND 338 342
FT TURN 343 346
FT STRAND 347 351
SQ SEQUENCE 448 AA; 47567 MW; A0759060C8F9E6CA CRC64;
Query Match 10.6%; Score 99; DB 1; Length 448;
Best Local Similarity 25.3%; Pred. No. 0.84;
Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;
QY 10 TADVRGIDCHALADRIGLDEAEIARLSTGIDD-----DTMAALAAE----- 53
DB 41 TPIIRNGGE---LTNLLGNSETTLARNEESATADLTAAAVDTVAARAAENAGAAWEA 97
QY 54 ----QLFEATADAL-----VTDFYDHLESYERTQDLFANSTKTKVQKETAQAEYLLG 102
DB 98 AAADALAKAKADALKEFNKYGVSDYYKNL-----INNKTVEGKDLQAQVW-- 145
QY 103 LGRGEYDTEVAARQARIGKHIVGLGPOVILGAYTRYTYTGLLDALADDDVADRGEEAAA 162
DB 146 -----ESAKKARISEATD-----GLSDFLKSTQTPA---EDTVK 175
QY 163 AVDELVARFL 172
DB 176 SIELAEAKVL 185
RESULT 5
SPG2_STRSG

ID AC P19909; SPG2_STRSG STANDARD; PRT; 593 AA.
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Immunoglobulin G binding protein G precursor (IgG binding protein G).
GN SPG.
OS Streptococcus sp. (Lancefield group G).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G148;
RX MEDLINE=88029445; PubMed=3665928;
RA Olsson A., Eliasson M., Guss B., Nilsson B., Hellman U.,
RA Lindberg M., Uhlen M.;
RT "Structure and evolution of the repetitive gene encoding
RT streptococcal protein G.";
RL Eur. J. Biochem. 168:319-324(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GX7805;
RX MEDLINE=86300657; PubMed=3017704;
RA Guss B., Eliasson M., Olsson A., Uhlen M., Frej A.-K., Joernvall H.,
RA Flock J.-I., Lindberg M.;
RT "Structure of the IgG-binding regions of streptococcal protein G.";
RL EMBO J. 5:1567-1575(1986).
RN [4]
RP STRUCTURE BY NMR OF 371-427.
RC STRAIN=G148;
RX MEDLINE=98290449; PubMed=9628485;
RA Malakauskas S.M., Mayo S.L.;
RT "Design, structure and stability of a hyperthermophilic protein
RT variant.";
RL Nat. Struct. Biol. 5:470-475(1998).
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X06173; CAA29540.1; -;
DR EMBL; Y00428; CAA68489.1; -;
DR EMBL; X04015; CAA27638.1; -;
DR PIR; S00128; S00128.
DR PDB; 1GB4; 22-JUL-98.
DR PDB; 1FCC; 20-JUL-95.
DR PDB; 1FCL; 19-SEP-01.
DR PDB; 1FD6; 19-SEP-01.
DR PDB; 1GJS; 21-MAR-02.
DR PDB; 1GJT; 09-AUG-01.
DR PDB; 2IGG; 31-JAN-94.
DR InterPro; IPR002988; GA.
DR InterPro; IPR005877; Gpos_Y5IRK.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000724; IgG_bind_B.
DR Pfam; PF01468; GA; 3.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF01378; IgG_binding_B; 3.
DR


```

OX NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MHOM/BR/82 / Isolate BA-2;
RA MEDLINE=93133867; PubMed=8421715;
RX Burns J.M. Jr., Shreffler W.G., Benson D.R., Ghalib H.W., Badaro R.,
RA Read S.G.; characterization of a kinsin-related antigen of
RT Leishmania chagasi that detects specific antibody in African and
RA American visceral leishmaniasis."
RL Proc. Natl. Acad. Sci. U.S.A. 90:775-779(1993).
CC -!- DEVELOPMENTAL STAGE: PREDOMINANT IN AMASTIGOTES.
CC -!- SIMILARITY: Belongs to the kinsin-like protein family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L07879; AAA29254.1; -
CC PIR; A47334; A47334.
CC HSSP; P17119; 3KAR.
CC InterPro; IPR001752; kinesin_motor.
CC Pfam; PF00225; kinesin; 1.
CC PRINTS; PR00380; KINESINHEAVY.
CC SMART; SM00129; KISC; 1.
CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
CC Motor protein; Microtubule; ATP-binding; Coiled coil; Repeat.
KW DOMAIN 1 399 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 426 >955 COILED COIL (POTENTIAL).
FT NP_BIND 122 129 ATP (POTENTIAL).
FT DOMAIN 704 >955 7 X 39 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 704 742 1.
FT REPEAT 743 781 2.
FT REPEAT 782 820 3.
FT REPEAT 821 859 4.
FT REPEAT 860 898 5.
FT REPEAT 899 937 6.
FT REPEAT 938 >955 7 (PARTIAL).
FT NON_TER 955 955
SQ SEQUENCE 955 AA; 106168 MW; 8CA76815BE84C6E9 CRC64;

Query Match 9.5%; Score 89; DB 1; Length 955;
Best Local Similarity 29.3%; Pred. No. 12;
Matches 49; Conservative 19; Mismatches 69; Indels 30; Gaps 7;

QY 44 DDTMAALAEQPLFEATADALVT---DFYDHLESYERTQDLFANSTKVEQLKETQAEVL 100
DB 546 ESTVAQLREQRERERVALDQTHQKQLEAESSERTA---AERDQLQLQTELQSE-R 601
QY 101 LGLGRGEYDTE-YAAQRAIRIGKHVDVLGLGPDVYLGA-----YTRYITGLL----- 145
DB 502 TQLSQVTDRELTLDLQIQEYGETELARDVALCAQEMEARVHAAPVHLQTLLELAT 661
QY 146 ---DALADDVADRGEAAAVDELIV-----ARFLPMLKLTFDQOI 184
DB 562 EWEDALRRERALARERDEAAAEALDAASTQNAESACERLTSLEQOL 708

RESULT 8
ID PLSB MYCLE STANDARD; PRT; 775 AA.
AC Q9X7B0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glyceral-3-phosphate acyltransferase (EC 2.3.1.15) (GPAT).
GN PLSB OR ML1246 OR MLCB1610.07.
OS Mycobacterium leprae.

```

```

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OX Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN NCBI_TaxID=1769;
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Biglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K.M., Devlin K., Duthoy S., Feltingwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RA "Massive gene decay in the leprosy bacillus."
RT Nature 409:1007-1011(2001).
CC -!- CATALYTIC ACTIVITY: Acyl-CoA + sn-glycerol 3-phosphate = CoA + 1-
CC acyl-sn-glycerol 3-phosphate.
CC -!- PATHWAY: De novo phospholipid biosynthesis; first step. May also
CC function in the regulation of membrane biogenesis.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
CC -!- SIMILARITY: Belongs to the GPAT / DAPAT family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL049913; CAB43153.1; -
CC PIR; AL583921; CAC31627.1; -
CC PIR; T45238; T45238.
CC Leproma; ML1246; -.
CC HAMAP; MF_00393; -. 1.
CC InterPro; IPR002123; Acyltransferase.
CC Pfam; PF01553; Acyltransferase; 1.
CC SMART; SM00563; PISC; 1.
CC Phospholipid biosynthesis; Transferase; Acyltransferase; Membrane;
KW Complete proteome.
SQ SEQUENCE 775 AA; 87363 MW; 03DD77C778293CDF CRC64;

Query Match 9.4%; Score 88; DB 1; Length 775;
Best Local Similarity 29.4%; Pred. No. 12;
Matches 40; Conservative 20; Mismatches 40; Indels 36; Gaps 7;

QY 32 EIAWRLSF---TGIDDDTMAALAEQPLFEATADALVTDFYDHLESYERTQDLFANSTK 87
DB 631 EIAWQNHEDRVSGDGDIDAMLLTKRPLI---SDAMLRVFF---EAYDIVADVLRDAPA 684
QY 98 TVEQKETQAEYLLGLGRGEYDTEYAAQ-RARIGKHVDVLGLGPDVYLGAIFYTRYITGLLD 146
DB 685 DVQGKELT-ELALGVGR-----QYVAQGRVRSGE-----SVSTLIF 719
QY 147 ALADDVADRGEAA 162
DB 720 ATAYQVVDQNLIAPA 735

RESULT 9
ID TIG MYCLE STANDARD; PRT; 469 AA.
AC Q9CBY2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trigger factor (TF).
GN TIG OR ML1481.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

```


DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Gas vesicle protein C.
 GN GVPC.
 OS Halobacterium mediterranei (Haloflex mediterranei).
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Haloflex.
 OX NCBI_TaxID=2252;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1411;
 RX MEDLINE=93021102; PubMed=1404376;
 RA Engler C., Krueger K., Offner S., Pfeifer F.;
 RT "Three different but related gene clusters encoding gas vesicles in
 RT halophilic archaea."
 RL J. Mol. Biol. 227:586-592 (1992).
 CC -!- FUNCTION: May confer stability to the gas vesicle membranes. Gas
 CC vesicles are small, hollow, gas filled protein structures that are
 CC found in several microbial planktonic microorganisms. They allow
 CC the positioning of the organism at the favorable depth for growth.
 CC -!- SUBCELLULAR LOCATION: Binds to the external surface of the gas
 CC vesicle membrane.
 CC -!- SIMILARITY: VERY DISTANT, TO GAS VESICLE PROTEINS TYPE C OF
 CC CYANOBACTERIA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X64701; CAA45944.1; -;
 DR PIR; S28115; S28115.
 DR InterPro; IPR008639; Halo GVPC.
 DR Pfam; PF05465; Halo GVPC; 1.
 KW Gas vesicle; Repeat.
 FT DOMAIN 22 274 7 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 22 59 1.
 FT REPEAT 60 84 2.
 FT REPEAT 85 122 3.
 FT REPEAT 123 160 4.
 FT REPEAT 161 192 5.
 FT REPEAT 193 232 6.
 FT REPEAT 233 274 7.
 FT DOMAIN 246 335 ASP/GLU-RICH (ACIDIC).
 SQ SEQUENCE 381 AA; 42653 MW; 9FB48199D0305921 CRC64;
 Query Match 8.8%; Score 82.5; DB 1; Length 381;
 Best Local Similarity 25.9%; Pred. No. 15;
 Matches 45; Conservative 27; Mismatches 73; Indels 29; Gaps 9;
 QY 11 ADVNRIGDHALRIGLDEARINRLSFTGIDDDTMAALAAEQPLFEATADALVTDFYD 70
 DB 81 ADRRDAFD--AYADIFATDVAEQDVSDLLAIDDLRAEMDTEAFDAYADAFYD--- 135
 QY 71 HLESYERTQDLPANSTKTKVEQLKQAEYLLGLGR-GEYDTYAAQARIGKHIVLGLG 129
 DB 136 -VATLRDVSGL-----LTAISELQSEFVSQGEFNGYSEFGAD---IDQFHAVAEK 184
 QY 130 PDVY---LQAYTRY---YTLG-LDALDDVVADRG-----EAAAADVELVARF 171
 DB 185 RDGHKDVADAFIQYREEFHGVQSLDNIIAAFQEMGYRKAFETTEAFASF 238
 RESULT 12
 VAL_BPSP STANDARD; PRT; 331 AA.
 AC P09677;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Readthrough protein A1 [Contains: Coat protein].
 OS Bacteriophage SP.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Leviviridae;
 OC Alloviridae.
 OX NCBI_TaxID=12027;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88289362; PubMed=3399390;
 RA Hirashima A., Hirose T., Inayama S., Inokuchi Y., Jacobson A.B.;
 RT "Analysis of the complete nucleotide sequence of the group IV RNA
 RT coliphage SP."
 RL Nucleic Acids Res. 16:6205-6221 (1988).
 CC -!- FUNCTION: Not yet known.
 CC -!- MISCELLANEOUS: The readthrough protein A1 includes the coat
 CC protein sequence.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X07489; CAB37299.1; -;
 DR PIR; S01964; S01964.
 DR HSP; P03615; 1QBE.
 DR InterPro; IPR002703; Levi_coat.
 DR Pfam; PF01819; Levi_coat; 1.
 KW Coat protein.
 FT CHAIN 1 132 COAT PROTEIN.
 FT CHAIN 1 331 READTHROUGH PROTEIN A1.
 SQ SEQUENCE 331 AA; 36203 MW; 4A6642E4B52C6582 CRC64;
 Query Match 8.8%; Score 82; DB 1; Length 331;
 Best Local Similarity 24.5%; Pred. No. 14;
 Matches 39; Conservative 17; Mismatches 61; Indels 42; Gaps 6;
 QY 37 LSFTGIDDDTMAALAAEQPLFEATADALVTDFYDHLESYERTQDLFANS-----T 86
 DB 94 LSFTSYSTDEERALIRTS-LAALLADPLIVDAIDNLNPAYWALLVASSGGGNDSPDV 152
 QY 87 KTVQLKETQAEYLLGLGRGEYDTYAAQARIGKHIVLGLG-PDVYLGAYTRYTGLL 145
 DB 153 PVVPVKPPD-----GTGRYKCPACY--RLGSIYEVKSGSDIY----- 191
 QY 146 DALADDVVADRGEEAAAADVELVARFLPMLKLLTDFDQI 184
 DB 192 -----ERGDEVSVTFDVALEDFLGNTRNRWDQL 221
 RESULT 13
 PROB_PSEAE STANDARD; PRT; 372 AA.
 AC Q9HVL9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glutamate 5-kinase (EC 2.7.2.11) (Gamma-glutamyl kinase) (GK).
 GN PROB OR P44565.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PAO1;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

```

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964 (2000).
CC -!- FUNCTION: Catalyzes the transfer of a phosphate group to glutamate
CC to form glutamate 5-phosphate which rapidly cyclizes to 5-
CC oxophosphate.
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate = ADP + L-glutamate 5-
CC phosphate.
CC -!- PATHWAY: Proline biosynthesis; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the glutamate 5-kinase family.
CC -!- SIMILARITY: Contains 1 PUA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE004870; AAG07953.1; -.
CC PIR: G83074; G83074.
CC HAMAP: MF_00456; -. 1.
CC InterPro: IPR001048; Aa_kinase.
CC InterPro: IPR001057; Glu_5kinase.
CC InterPro: IPR005715; ProB.
CC InterPro: IPR002478; PUA.
CC Pfam: PF00696; aakkinase; 1.
CC Pfam: PF01472; PUA; 1.
CC PRINTS: PR00474; GLU5KINASE.
CC SMART: SM00359; PUA; 1.
CC TIGRFAMs: TIGR01027; proB; 1.
CC PROSITE: PS00902; GLUTAMATE_5_KINASE; 1.
CC PROSITE: PS00890; PUA; 1.
CC Transferase; Kinase; Proline biosynthesis; Complete proteome.
KW DOMAIN 280 358
FT SEQUENCE 372 AA; 39845 MW; FAF2B91F6A8DEC36 CRC64;
SQ
Query Match 8.8%; Score 82; DB 1; Length 372;
Best Local Similarity 25.3%; Pred. No. 16;
Matches 49; Conservative 19; Mismatches 66; Indels 60; Gaps 8;
QY 3 NDNDLTATDVNRGIDGHADRLGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATAD 62
Db 138 NENDTVVTVDIEIRFG-----DNDTLAALVAN--LVEADLL 169
QY 63 ALVTD---FYDHLSEYERTQDLFANSTKTVEQLKETAQYLLGLGRGEYDTEYAAQR--A 117
Db 170 VILTDROGMEDADPRNPNDAQLIYEARADDPQLDAVAGGSAGALRGGMQTKLRAALAA 229
QY 118 RIGKIHVDVLGLGPDVYLGAITYRYTGLLDALADDWA-----DRGEAAAAVDEL 167
Db 230 RSG-----GHTVIVG-----GRIERVLDRLRAGERLGLTLTPDRSEKAR--KQW 272
QY 168 VARFLPMLKILTFD 181
Db 273 LAGHLQMRGTLVLVD 286
RESULT 14
IF2_SYNY3
ID IF2_SYNY3 STANDARD; PRT; 1001 AA.
AC P72689;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Translation initiation factor IF-2.
GN INFB OR SLR0744.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

```

```

OX NCBI_TaxID=11148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the IF-2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D90900; BAA16696.1; -.
CC PIR: S74544; S74544.
CC HAMAP: MF_00100; -. 1.
CC InterPro: IPR000795; EF_GTPbind.
CC InterPro: IPR004161; EFTU_D2.
CC InterPro: IPR000178; IF2.
CC InterPro: IPR006847; IF2_N.
CC InterPro: IPR005225; Small_GTP.
CC InterPro: IPR009000; Translat_factor.
CC Pfam: PF00009; GTP_EFTU; 1.
CC Pfam: PF03144; GTP_EFTU_D2; 2.
CC Pfam: PF04760; IF2_N; 2.
CC PRINTS: PR00315; ELONGATNFACT.
CC PRODOM: PD186100; IF2; 1.
CC TIGRFAMs: TIGR00487; IF-2; 1.
CC TIGRFAMs: TIGR00231; small_GTP; 1.
CC PROSITE: PS01176; IF2; 1.
CC Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 496 648
FT NP_BIND 502 509 GTP (BY SIMILARITY).
FT NP_BIND 552 556 GTP (BY SIMILARITY).
FT NP_BIND 606 609 GTP (BY SIMILARITY).
SQ SEQUENCE 1001 AA; 108118 MW; 4C7F00DEE56CE31C CRC64;
Query Match 8.7%; Score 81; DB 1; Length 1001;
Best Local Similarity 22.5%; Pred. No. 58;
Matches 47; Conservative 20; Mismatches 54; Indels 88; Gaps 7;
QY 22 LADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVDFVHLESYERTQDL 81
Db 423 LADLLKISETDIIRKLFKGV-----AVQITQTLDEETARMVAESFEVAVETPRV--- 473
QY 82 FANSTKTVEQLKETAQYLL-----GLGR-- 105
Db 474 -AAAAKTTEMLDEADLNLVRRPPVVTIMGHVDHGKTTLLDSIRKTKVAQGEAGGIQHI 532
QY 106 GEYDTE-----YAAOPARIKGIHVDVLGLGPDVYLGAITYRYTGLL 145
Db 533 GAYHVEVHNKTEQIVFLDTPGHEAFTAMRGAQKVTDI-----AIL 575
QY 146 DALADDVWADRGEA-----AAAVDELVA 169
Db 576 VVAADGVQPTKEAISHAKAAGVPLIVA 604

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:22:05 ; Search time 22.0533 Seconds
(without alignments)
2632.505 Million cell updates/sec

Title: US-09-455-978b-77

Perfect score: 933

Sequence: 1 MSNDNDTLVTADVRNGIDGH.....DELVARFLPKLLTFDDQI 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_verticillate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	148	15.9	439	16	Q9KFH4	Q9KFH4 bacillus ha
2	132	14.1	537	16	Q9AB06	Q9AB06 caulobacter
3	117.5	12.6	499	16	Q8UGU0	Q8UGU0 agrobacteri
4	105.5	11.3	555	16	Q9A5Y0	Q9A5Y0 caulobacter
5	102	10.9	434	16	Q814K6	Q814K6 bacillus ce
6	98.5	10.6	1222	16	Q8D4B1	Q8D4B1 vibrio vuln
7	97.5	10.5	248	10	Q8S4W2	Q8S4W2 phytophthor
8	97	10.4	218	16	Q88031	Q88031 streptomyce
9	95.5	10.2	1364	2	Q84GK0	Q84GK0 escherichia
10	95	10.2	439	16	Q81JN0	Q81JN0 bacillus an
11	94.5	10.1	419	9	Q8W6U7	Q8W6U7 bacterioph
12	94.5	10.1	883	17	Q9HRW3	Q9HRW3 halobacteri
13	94	10.1	218	16	Q82MB8	Q82MB8 streptomyce
14	94	10.1	1090	10	O64476	O64476 arabidopsis
15	93.5	10.0	514	16	Q8FUX2	Q8FUX2 brucella su
16	93	10.0	881	16	Q8UUA1	Q8UUA1 agrobacteri

17	92.5	9.9	503	16	Q8YDG6	Q8YDG6 brucella me
18	92.5	9.9	514	2	Q8VOK3	Q8VOK3 brucella ab
19	90.5	9.7	194	16	Q8DHH0	Q8DHH0 synchococc
20	89.5	9.6	400	16	Q82AV2	Q82AV2 streptomyce
21	88.5	9.5	458	16	Q8PMR9	Q8PMR9 xanthomonas
22	88.5	9.5	505	11	Q9KK9	Q9KK9 mus musculu
23	88.5	9.5	1361	16	Q8RHV1	Q8RHV1 fusobacteri
24	88	9.4	500	16	Q9JYD3	Q9JYD3 neisseria m
25	88	9.4	505	16	Q9UTB6	Q9UTB6 neisseria m
26	87.5	9.4	2246	2	Q9AKS6	Q9AKS6 pseudomonas
27	86.5	9.3	157	17	Q9HNO8	Q9HNO8 halobacteri
28	86.5	9.3	468	16	Q8P6N2	Q8P6N2 xanthomonas
29	86.5	9.3	895	11	Q8CL17	Q8CL17 mus musculu
30	86	9.2	670	4	Q96CNS	Q96CNS homo sapien
31	86	9.2	508	16	Q8CJX8	Q8CJX8 streptomyce
32	85.5	9.2	965	16	Q81YS7	Q81YS7 bacillus an
33	85	9.1	306	16	Q8D7S6	Q8D7S6 vibrio vuln
34	85	9.1	2703	16	Q9KOT0	Q9KOT0 neisseria m
35	84.5	9.1	266	17	Q9HND3	Q9HND3 halobacteri
36	84.5	9.1	759	16	Q98C20	Q98C20 rhizobium l
37	84	9.0	170	3	Q871H7	Q871H7 neurospora
38	84	9.0	412	16	Q7UG88	Q7UG88 rhodospirell
39	84	9.0	421	2	Q9RQ25	Q9RQ25 amycolatops
40	84	9.0	467	16	Q87DW5	Q87DW5 xylella fas
41	84	9.0	491	16	Q92TQ4	Q92TQ4 rhizobium m
42	83.5	8.9	258	16	Q82GU5	Q82GU5 streptomyce
43	83.5	8.9	712	10	Q80890	Q80890 arabidopsis
44	83.5	8.9	755	10	Q8RX83	Q8RX83 arabidopsis
45	83.5	8.9	826	16	Q92TC2	Q92TC2 rhizobium m

ALIGNMENTS

RESULT 1

Q9KFH4	PRELIMINARY;	PRT;	439 AA.
ID	Q9KFH4		
AC	Q9KFH4;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	Methyl-accepting chemotaxis protein.		
GN	BH0505		
OS	Bacillus halodurans.		
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
OX	NCBI_TaxID=86665;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C-125 / JCM 9153;		
RX	MEDLINE=20512582; PubMed=11058132;		
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,		
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,		
RA	Horikoshi K.;		
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus		
RT	halodurans and genomic sequence comparison with Bacillus subtilis.";		
RL	Nucleic Acids Res. 28:4317-4331 (2000).		
DR	EMBL; AP001508; BAB04224.1; -		
DR	PIR; A83713; A83713.		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0004871; F:signal transducer activity; IEA.		
DR	GO; GO:0005351; F:sugar porter activity; IEA.		
DR	GO; GO:0006935; P:chemotaxis; IEA.		
DR	GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. .; IEA.		
DR	GO; GO:0007165; P:signal transduction; IEA.		
DR	InterPro; IPR002114; HPr SerP_S.		
DR	Pfam; PF00015; MCPsignal; 1.		
DR	SMART; SM00283; MA; 1.		
DR	PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.		
DR	PROSITE; PS00589; PTS_HPR_SER; 1.		
KW	Complete proteome.		
SQ	SEQUENCE 439 AA; 48918 MW; 1E092BB2F42592AE CRC64;		

Query Match	15.9%	Score 148;	DB 16;	Length 439;
Best Local Similarity	25.2%;	Pred. No. 0.00015;		
Matches 39;	Conservative 35;	Mismatches 71;	Indels 10;	Gaps 3

Qy	30	E	A	I	A	M	F	L	T	G	I	D	D	D	T	M	A	A	A	E	O	P	L	E	A	T	A	D	A	L	V	T	D	F	D	H	L	S	S	Y	E	R	T	Q	D	L	F	A	N	S	T	K	T	V	89
Db	33	E	S	E	L	S	A	O	R	M	I	H	L	T	D	L	K	R	M	A	L	O	P	L	V	E	N	N	E	V	L	A	D	A	F	Y	S	N	I	K	O	P	N	E	I	-	E	T	H	S	V	91			
Qy	90	E	O	I	K	E	T	O	A	E	Y	L	L	G	R	G	E	Y	T	E	Y	A	A	O	R	I	G	K	I	H	D	V	L	G	L	G	P	D	V	I	G	A	T	R	Y	T	T	G	L	L	D	A	149		
Db	92	E	R	L	K	E	T	L	K	O	H	I	E	M	F	N	G	E	I	D	O	A	F	L	T	Q	K	L	Q	I	A	A	H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	143						
Qy	150	D	O	V	V	A	D	R	G	E	A	A	A	A	A	V	D	E	L	V	A	R	F	L	P	M	L	K	L	L	T	D	Q	Q	I	184																			
Db	144	D	S	L	I	-	O	L	L	E	O	H	I	O	S	P	S	D	I	V	A	T	R	S	L	K	L	N	L	E	O	O	L	177																					

RESULT 2

ID	Q9AB06	PRELIMINARY;	PRT;	537 AA.
AC	Q9AB06;			
DC	01-JUN-2001	(TrEMBLrel. 17, Created)		
DT	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)		
DT	01-JUN-2003	(TrEMBLrel. 24, Last annotation update)		
DE	DE	Methyl-accepting chemotaxis protein McpB.		
GN	CC0428.			
OS	Caulobacter crescentus.			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;			
OC	Caulobacteraceae; Caulobacter.			
OX	NCBI_TaxID=155892;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 19089 / CB15;			
RC	MDRLINE=2173698; PubMed=11259647;			
RX	Nierman W.C., Feldberg T.V., Laub M.T., Paulsen I.T., Nelson K.E.,			
RA	Eisen J., Heidelberg J.F., Alley M.R.K., Onta N., Maddock J.R.,			
RA	Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,			
RA	DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,			
RA	Kolony J.P., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,			
RA	Uterback T., Tran K., Wolf A., Vamathavan J., Emolaeva M., White O.,			
RA	Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;			
RT	"Complete genome sequence of <i>Caulobacter crescentus</i> ."			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).			

```

DR EMBL; AE005715; AAK22415.1; -.
DR PIR; C87302; C87302.
DR HSP; P02942; 1QU7.
DR TIGR; CC0428; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_chemotaxis.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 1.
DR Complete proteome.
DR SO SEQUENCE 537 AA; 57374 MW; 1B4B797C52D5DDE1 CRC64;

```

Query Match 14.1%; Score 132; DB 16; Length 537;
Best Local Similarity 27.2%; Pred. No. 0.0052;
Matches 41; Conservative 26; Mismatches 72; Indels 12; Gaps 3;

Qy	33	I	A	R	L	S	F	G	T	G	I	D	D	T	M	A	L	A	A	B	O	P	E	A	T	A	D	A	L	V	T	O	F	Y	D	H	L	E	S	Y	E	R	T	Q	D	I	F	A	N	S	T	K	T	V	E	Q	L	92			
Db	11	I	C	E	R	T	A	F	M	G	I	D	D	K	A	R	S	A	L	R	D	L	R	P	V	I	R	A	E	I	G	K	A	L	D	N	F	Y	G	K	V	R	A	T	P	E	T	R	K	F	F	S	D	D	R	H	N	A	A	S	70
Qy	93	K	E	T	O	A	E	I	L	L	G	K	G	B	Y	D	T	E	A	A	Q	A	R	I	G	K	H	D	V	L	G	I	G	P	D	V	I	G	A	Y	T	R	Y	T	T	G	L	L	D	A	L	A	D	D	V	152					

Db	71	SRQAHGCV-LAEQFSDDDYQAVRAITGCTHARIGLEPRWYIGGVAVVGDHLVRAVIDSM	129
Qy	153	-----VADRGEEAAAAVDELVARFLPM	174
Db	130	WPRGILAKGGSDRAGEAVAL--MKAIFLDM	158

RESULT 3

Q8UGLO
ID Q8UGLO
PRELIMINARY;
PRT: 499 AA.

AC Q6UJ01; 2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DT 01-JUN-2002 (TREMBLrel. 24, Last annotation update)
 DE Methyl-accepting chemotaxis protein.
 DE MCVB OR AFU1027 OR AGR_C1888.
 GN MCVB
 GN Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OS Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 NCBI TaxID=176299;
 OX NCBI TaxID=176299;

[1] —
 RN SEQUENCE FROM N.A.
 RP MEDLINE=21608550; PubMed=11743193;
 RX Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McAlleland E., Palmeri A.,
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Doan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT The genome of the natural genetic engineer *Agrobacterium tumefaciens*
 RT C58.";
 RL Science 294:2317-2323 (2001) .

science 294:2317-2323 (2001).

[2] SEQUENCE FROM N.A.
RX MEDLINE=21608554; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurrello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Muller
RA Houniel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markel
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.
RT "Genome sequence of the plant pathogen and biotechnology agent

RI genome sequenc... P C58. ",
RL Science 294:2323-2328 (2001).
DR EMBL; AE009067; AAL42040.1; -.
DR EMBL; AE008033; AAK86834.1; -.
DR PIR; A97485; A97485.
DR PIR; AB2703; AB2703.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007135; F:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_chemotaxis.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUC.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; NA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 1.
DR Complete proteome.
KW SEQUENCE 499 AA; 53808 MW; 9644888637D68ECE CRC64;

Query Match 12.6%; Score 117.5; DB 16; Length 499;
Best Local Similarity 26.5%; Pred. No. 0.088;
Matches 35: Conservative 23; Mismatches 67; Indels 7;

Qv 19 GHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDIFYDHLSEYRT 78

```

Db      3  GOAKTR-OLDE-----RLNFLGLGHERQNLSDMKGVITGSLDASLDREYTKVRVPET 56
QY      79  QDLFANSTKVEQLKETOAYLLGLGRGEYDTYAAQARIGIKHVDVLGPDVYLGAYT 138
Db      57  AKFF-SSEAHIIHAKSMQLKHSRIASGTFNEDYTNVTAIGRTHARLGLRPRWYIGYA 115
QY      139  RYTGLLDALAD 150
Db      116  LMLDGIVKAVIE 127

RESULT 4
Q9A5Y0
ID      Q9A5Y0      PRELIMINARY;      PRT;      555 AA.
AC      Q9A5Y0;
DT      01-JUN-2001 (TrEMBLrel. 17, Created)
DT      01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT      01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE      Methyl-accepting chemotaxis protein McpM.
GN      CC2317.
OS      Caulobacter crescentus.
OC      Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC      Caulobacteraceae; Caulobacter.
OX      NCBI_TaxID=155892;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=21173698; PubMed=11259647;
RA      Niernan W.C.; Feldblyum T.V.; Laub M.T.; Paulsen I.T.; Nelson K.E.,
RA      Eitsen J.; Heidelberg J.F.; Alley M.R.K.; Ohta N.; Maddock J.R.,
RA      Potocka I.; Nelson W.C.; Newton A.; Stephens C.; Phadke N.D.; Ely B.,
RA      DeBoy R.T.; Dodson R.J.; Durkin A.S.; Gwinn M.L.; Haft D.H.,
RA      Kolonay J.F.; Smit J.; Craven M.B.; Khouri H.; Shetty J.; Berry K.,
RA      Utterback T.; Tran K.; Wolf A.; Vamathevan J.; Ermolaeva M.; White O.,
RA      Salzberg S.L.; Venter J.C.; Shapiro L.; Fraser C.M.;
RT      "Complete genome sequence of Caulobacter crescentus.";
RL      Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR      EMBL; AE005901; AAK24288.1; -.
DR      PIR; D87536; D87536.
DR      HSSP; P02942; LQUT.
DR      TIGR; CC2317; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0004871; F:signal transducer activity; IEA.
DR      GO; GO:0006935; P:chemotaxis; IEA.
DR      GO; GO:0007165; P:signal transduction; IEA.
DR      InterPro; IPR004089; Chntaxis_transd.
DR      Pfam; PF00015; MCPsignal; 1.
DR      SMART; SM00304; HAMP; 1.
DR      SMART; SM00283; MA; 1.
DR      PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR      PROSITE; PS50885; HAMP; 1.
KW      Complete proteome.
SQ      SEQUENCE 555 AA; 59114 MW; C9560265C1EC0B31 CRC64;

Query Match      11.3%; Score 105.5; DB 16; Length 555;
Best Local Similarity 24.7%; Pred. No. 1.2;
Matches 40; Conservative 28; Mismatches 77; Indels 17; Gaps 5;

QY      31  AEIAWRLSFTGIDDDTMAALAAEQFLFEATADALVDFYDHLESYERTQDLFANSTKIVE 90
Db      46  AKLDQMAFMRFDSRAHLRAIKPVIDAIEGAALGFQYSQVRLFPDTRVKFRDDGH-MA 104
QY      91  QLKETOAYLLGLGRGEYDTYAAQARIGIKHVDVLGPDVYLGAYTRYTGLLDALAD 150
Db      105  GAERAAQAHWRIAPAGGESYVRDVERIGRSHADADIAPOWYIGGYAVVVEVMRL-- 162
QY      151  DWADRG-----EAAAAVDELV-ARFLPM-LKLLTF 180
Db      163  --VAKRAGLFSNAKSDAELADGLSALIKAAFLMDLSVSTY 202

```

```

RESULT 5
Q814K6
ID      Q814K6      PRELIMINARY;      PRT;      434 AA.
AC      Q814K6;
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Methyl-accepting chemotaxis protein.
GN      BC5424.
OS      Bacillus cereus (strain ATCC 14579 / DSM 31).
OC      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX      NCBI_TaxID=226900;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22608415; PubMed=12721630;
RA      Ivanova N.; Sorokin A.; Anderson I.; Galleron N.; Candelson B.,
RA      Kapatal V.; Bhattacharyya A.; Reznik G.; Mikhailova N.; Lapidus A.,
RA      Chu L.; Mazur M.; Goltsman E.; Larsen N.; D'Souza M.; Walunas T.,
RA      Grechkin Y.; Pusch G.; Haselkorn R.; Fonstein M.; Ehrlich S.D.,
RA      Overbeek R.; Kyrpides N.;
RT      "Genome sequence of Bacillus cereus and comparative analysis with
RT      Bacillus anthracis.";
RL      Nature 423:87-91(2003).
DR      EMBL; AE017015; AAP12286.1; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0004871; F:signal transducer activity; IEA.
DR      GO; GO:0006935; P:chemotaxis; IEA.
DR      GO; GO:0007165; P:signal transduction; IEA.
DR      InterPro; IPR004089; Chntaxis_transd.
DR      Pfam; PF00015; MCPsignal; 1.
DR      PRINTS; PR00260; CHEMTRNSDUCR.
DR      SMART; SM00283; MA; 1.
DR      PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
KW      Complete proteome.
SQ      SEQUENCE 434 AA; 49441 MW; E1007BA0230620C5 CRC64;

Query Match      10.9%; Score 102; DB 16; Length 434;
Best Local Similarity 20.5%; Pred. No. 1.7;
Matches 34; Conservative 27; Mismatches 71; Indels 34; Gaps 3;

QY      31  AEIAWRLSFTGIDDDTMAALAAEQFLFEATADALVDFYDHLESYERTQDLFANSTK--- 87
Db      34  SELKIQMDMLHISKEDLQVVKVQFPFVAEIDWITEKFY-----ANITKQPN 80
QY      88  -----TVEQLKETOAYLLGLGRGEYDTYAAQARIGIKHVDVLGPDVYLGAYT 138
Db      81  LITTIERYSSIPKLKQTLKTHIKELFSGNMHEDFTEQVRIAKRVQIGLHKWYTAAYQ 140
QY      139  RYTGLLDALADDVVADRGEAAAVDELVARFLPMLKLLTFDQOI 184
Db      141  ELFRSIKILQTKI-----TTIDDFSYSINVKLFTLEQEL 177

RESULT 6
Q8D4B1
ID      Q8D4B1      PRELIMINARY;      PRT;      1222 AA.
AC      Q8D4B1;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Conserved hypothetical protein.
GN      VV21397.
OS      Vibrio vulnificus.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC      Vibrionaceae; Vibrio.
OX      NCBI_TaxID=672;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STEAIN=CMCP6;
RA      Rhee J.H.; Kim S.Y.; Chung S.S.; Kim J.J.; Moon Y.H.; Jeong H.,
RA      Choy H.E.;
RT      "Complete genome sequence of Vibrio vulnificus CMCP6.";

```

RL	Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AE016813; AAC08278.1; -	
DR	InterPro; IPR000437; Prok lipoprot S.	
DR	PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.	
KW	Hypothetical protein; Complete proteome.	
SQ	SEQUENCE 1222 AA; 136974 MW; B04525D301FA99E2 CRC64;	
	Query Match 10.6%; Score 98.5; DB 16; Length 1222;	
	Best Local Similarity 20.2%; Pred. No. 15;	
	Matches 37; Conservative 34; Mismatches 59; Indels 53; Gaps 7;	
QY	43 DDDTMAALAAEPLFEATADALVTFDYHLESY-----ERTQDLFANSTKTIV 89	
DB	606 DSENTYRLTEHOPHPIWSSPTDSSYEBYADY--NYNQKWNENTAYVWVTELLADSSIKV 663	
QY	90 -----EQKKEQTQAE-----YLLGLGRGEYDT-----EYAAQ 115	
DB	664 SGRVYLADSDAIKETSSEKPGKPFSESLHERTISATEKRAYETWHHIALEGYDFTAS 723	
QY	116 RARIQKIHVDVLGLGPDVYLG-AYTRYTGLLDALADDVVADRGEEAAAADVELVARFLPM 174	
DB	724 ANVFGHQYTVFNLYDIIWGYSTQWFDNRNIDLIAE---VDRLLGQGVVDQITPEMIEN 780	
QY	175 LKL 177	
DB	781 LNL 783	
	RESULT 7	
ID	Q8S4W2 PRELIMINARY; PRT; 248 AA.	
AC	Q8S4W2;	
DT	01-JUN-2002 (TrEMBLrel. 21, Created)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	Syntaxin 6.	
OS	Phytophthora infestans (Potato late blight fungus).	
OC	Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;	
OC	Phytophthora.	
OX	NCBI_TaxID=4787;	
[1]	SEQUENCE FROM N.A.	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21947832; PubMed=11950882;	
RA	Dacks J.B., Doolittle W.F.;	
RT	"Novel syntaxin gene sequences from Giardia, Trypanosoma and algae:	
RT	implications for the ancient evolution of the eukaryotic endomembrane	
RT	system.";	
RL	J. Cell Sci. 115:1635-1642(2002).	
DR	EMBL; AF404749; AM12665.1; -	
DR	InterPro; IPR000727; T_SNARE.	
DR	Pfam; PF05739; SNARE; 1.	
DR	SMART; SM00397; T_SNARE; 1.	
DR	PROSITE; PS50192; T_SNARE; 1.	
SQ	SEQUENCE 248 AA; 27934 MW; 2FE6891BA5AA0067 CRC64;	
	Query Match 10.5%; Score 97.5; DB 10; Length 248;	
	Best Local Similarity 22.7%; Pred. No. 1.9;	
	Matches 42; Conservative 31; Mismatches 67; Indels 45; Gaps 7;	
QY	6 DTLVTADVNRGIDGHALADRLGLDEAEIAWRLSF-----TGIDDDTMAA--LA 51	
DB	68 ETIIVWEANRAKEEH-----IDAEIASRKAFVAATRKELQAVSTEISTDVTKTRIK 120	
QY	52 AEQPLFEATADALVTFDYHLESYERTQDLFANSTKTVEQLKTOAYLLGLGRGEYDTE 111	
DB	121 EERKLMPQAKSS--TSFRNLTGQERNERFLEDETQROQQIQMFQNDISLAGL----- 170	
QY	112 YAAQRAPIKIHVDVLGLGPDVYLGAYTRYTGLLDALADDVVADRGEEAAAADVELVARF 171	
DB	171 ----HSDITRLH-----GVTVESISSEVKQNKMLDLDLTDVV-----DEAQERWNVWGRL 216	
QY	172 LPLMK 176	
	RESULT 8	
ID	O88031 PRELIMINARY; PRT; 218 AA.	
AC	O88031;	
DT	01-NOV-1998 (TrEMBLrel. 08, Created)	
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	Hypothetical protein SC06675.	
GN	SC06675 OR SC5A7.25C.	
OS	Streptomyces coelicolor.	
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
OC	Streptomycineae; Streptomycetaceae; Streptomyces.	
OX	NCBI_TaxID=1902;	
[1]	SEQUENCE FROM N.A.	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=A3(2);	
RA	Seeger K.J., Harris D.;	
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.	
[2]	SEQUENCE FROM N.A.	
RC	STRAIN=A3(2);	
RA	Parkhill J., Barrell B.G., Rajandream M.A.;	
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.	
[3]	SEQUENCE FROM N.A.	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=A3(2);	
RX	MEDLINE=97000351; PubMed=8843436;	
RA	Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,	
RA	Kinashi H., Hopwood D.A.;	
RT	"A set of ordered cosmids and a detailed genetic and physical map for	
RT	the 8 Mb Streptomyces coelicolor A3(2) chromosome.";	
RL	Mol. Microbiol. 21:77-96(1996).	
[4]	SEQUENCE FROM N.A.	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=A3(2) / M145;	
RX	MEDLINE=21996410; PubMed=12000953;	
RA	Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,	
RA	Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,	
RA	Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,	
RA	Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,	
RA	Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,	
RA	Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,	
RA	Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,	
RA	Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,	
RA	Hopwood D.A.;	
RT	"Complete genome sequence of the model actinomycete Streptomyces	
RT	coelicolor A3(2).";	
RL	Nature 417:141-147(2002).	
DR	EMBL; AL939128; CAAL19954.1; -	
DR	PIR; T35174; T35174.	
DR	GO; GO:0003677; F:DNA binding; IEA.	
DR	InterPro; IPR001387; HTH_3.	
DR	Pfam; PF01381; HTH 3; 1.	
DR	SMART; SM00530; HTH XFE; 1.	
KW	Hypothetical protein; Complete proteome.	
SQ	SEQUENCE 218 AA; 23670 MW; D22747AF816D1DE4 CRC64;	
	Query Match 10.4%; Score 97; DB 16; Length 218;	
	Best Local Similarity 29.6%; Pred. No. 1.7;	
	Matches 53; Conservative 21; Mismatches 77; Indels 28; Gaps 10;	
QY	7 TLVTADVNRGIDGHALADRLGLDEAEI--AWRLS-FTGIDDDTMAALAAEQLFEATADA 63	
DB	15 TLAADVAVRG-----ALADRLDVPHAEVDFVGRUSAAAGSPESVVGALLGRP---TGGA 66	
QY	64 LV-TDFVDHLESYERTQDLFANSTKTVEQ-----LKETQAYLLGLGRGEYDTEYAAQ 115	
DB	67 DVQTRFQRLDLRLRRF-LKPNCRKYTQEQEADGAGMSRQAGALIN-GDRRPTMEHCDA 124	
QY	116 RARIQKIH--VLGLGPDVYLGAYTRYTGLLDALADDVVADRGEEAAAADVELVARFL 172	


```
Db 125 IQRRFRVHAGFLTAEDPEALGALQRTQELLQQL-----ADRRGQAAPADDDPLERLL 178

RESULT 9
Q84GKO PRELIMINARY; PRT; 1364 AA.
AC Q84GKO;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Secreted autotransporter protein Bata.
GN BATA.
OS Escherichia coli.
OG Plasmid pCS1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H10407;
RA Fleckenstein J.M., Patel S.K., Dotson J.;
RT "Identification and molecular characterization of Bata, an
RT autotransporter protein of enterotoxigenic Escherichia coli.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY163491; AAC017297.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:004252; F:serine-type endopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR000710; Peptidase S6.
DR InterPro; IPR004899; Pertactin.
DR Pfam; PF02395; IGAL1; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR TIGRFAMs; TIGR01414; autotrans_barl; 1.
DR Plasmid.
KW Plasmid.
SQ SEQUENCE 1364 AA; 147696 MW; 33B341FDAB6859E2 CRC64;

Query Match 10.2%; Score 95.5; DB 2; Length 1364;
Best Local Similarity 25.5%; Pred. No. 32;
Matches 48; Conservative 28; Mismatches 83; Indels 29; Gaps 9;

Qy 5 NDTLVTADVNRGIDGH---ALADRIGLDEAETAWELSTGIDDDTMAALAAEQLPFATA 61
Db 1012 NVTLITAP--KGSDETFKAGTQQIGF--SNITPEIRTEITDTATQWYLTGYQSVADARA 1067
Qy 62 DALVTDFYDHLSEYRTQDLFANSTKTVEQLKETQ-----AEVLLGLGRGE--YDTEYA 113
Db 1068 SKIATDFMD--SGYKSFTEVNNLNKRMGLRDSQDAGGAWRMNGTSGESGYRDNVT 1125
Qy 114 AQRARIGKIHVDLGLGPDVYLGAATRY-----YTGLLDALADDVADVRGEEAAAV 164
Db 1126 HVQIGADRKHELNGI--DLFTGALLTYDNNASSQAFSGTKSLGGVYAGLPESGAYF 1183
Qy 165 DELVARFL 172
Db 1184 D-LIGKYL 1190

RESULT 10
Q81JUN0 PRELIMINARY; PRT; 433 AA.
AC Q81JUN0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN BA5673.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
```

```
OX NCBI_TaxID=198094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madupu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomason B., Friedlander A.M., Koshler I.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86(2003).
DR EMBL; AB017041; AAP29306.1; -.
DR TIGR; BA5673; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis trans.
DR InterPro; IPR004090; Me chemotaxis.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 433 AA; 49391 MW; E06512BA0E696D92 CRC64;

Query Match 10.2%; Score 95; DB 16; Length 433;
Best Local Similarity 19.3%; Pred. No. 6.9;
Matches 32; Conservative 29; Mismatches 71; Indels 34; Gaps 3;

Qy 31 AEIARWLSFTGIDDDTMAALAAEQLPFATAVDLVDYDHLSEYRTQDLFANSTK--- 87
Db 33 SELKVQMDMLHISKEDLQIVKVLQFFTYEEDITWTEKFF-----ANITKPN 79
Qy 88 -----TVEQLKETQAEYLLGLGRGEYDTEYAAQARIGKIHVDLGLGPDVYLGAAT 138
Db 80 LITIERYSSTPKLQTLKTHIKELFSGDMHEDFIEQRVKIARHVVQIGLHKRWYTAAYQ 139
Qy 139 RYTTGLLDALADDVADVRGEEAAAVDELVARFLPMLKLLTFDDQI 184
Db 140 ELFRSINKILTKI-----TTIDDFSYSINVKLFTLEQEL 176

RESULT 11
Q8W6U7 PRELIMINARY; PRT; 419 AA.
AC Q8W6U7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Gp4.
GN 4.
OS Bacteriophage phiE125.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda-like viruses.
OX NCBI_TaxID=180504;
RN [1]
RP SEQUENCE FROM N.A.
RA Woods D.E., Jeddleloh J.A., Fritz D.F., DeShazer D.;
RT "Burkholderia thailandensis E125 Harbors a Temperate Bacteriophage
RT Specific for Burkholderia mallei.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF447491; AAL40277.1; -.
DR InterPro; IPR006944; Phage portal.
DR InterPro; IPR006427; Portal_HK97.
DR Pfam; PF04860; Phage_portal; 1.
```

DR TTGRFMS; TTGR01537; portal_HK97; 1.
SQ SEQUENCE 419 AA; 46404 MW; C15C3CC08F21D2C1 CRC64;

Query Match 10.1%; Score 94.5; DB 9; Length 419;
Best Local Similarity 21.7%; Pred. No. 7.2;
Matches 49; Conservative 28; Mismatches 74; Indels 75; Gaps 10;

Qy 5 NDTLVTVADVNGIDGHALADRIGLDRAETIAWRLSFTGIDDDTWAALAAE-----QPLFEAT 60
Db 201 NGTALSGVIERPKDAPALKDQASVDRTDGNNAKFCGSGNAKKVALLQEGMTFRPLSMTN 260
Qy 61 ADALVTD-----FYDHL-----SYE 76
Db 261 VDAALIDALISALDIARIYKIPAHMVNELERATFSNIEHQSLQFVIYITLLPWKRHEQA 320
Qy 77 RTQDLFANSTKTVEQLKETOAEY-LIGLGRGEYDTEYAAQRARIGK-----IHDVLGI- 128
Db 321 KTRDLLLPSSR-----KQYEIYNLAGLLRGDQSSRYAAV-AVGROGWLMSINDIRLE 373
Qy 129 -----GPDVYLGAFTYTYTGLDLA-LADDVADRGEEAAAADDEL 167
Db 374 NMPVKGDIYLS-----NMVDAKPKQOLPVKSEPTKAAIDEI 414

RESULT 12
Q9HRW3 PRELIMINARY; PRT; 883 AA.
ID Q9HRW3
AC Q9HRW3; PRT; 883 AA.
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Vng0514C.
CN Vng0514C.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M., Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J., Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A., Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W., Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H., Isenbarger T.A., Beck R.F., Pohlschroder M., Spudich J.L., Jung K.-H., Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D., Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dasgupta S., Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dasgupta S., Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RL "Genome sequence of Halobacterium species NRC-1".
DR EMBL; AE005003; AAG19045.1; -.
DR PIR; A84210; A84210.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0007059; P:chromosome segregation; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF02463; SMC_N; 1.
DR PRINTS; PR00194; TROPOMYOSIN.
KW Complete proteome.
SQ SEQUENCE 883 AA; 96969 MW; F8284D3326C92BB9 CRC64;

Query Match 10.1%; Score 94.5; DB 17; Length 883;
Best Local Similarity 28.3%; Pred. No. 21;
Matches 51; Conservative 23; Mismatches 63; Indels 43; Gaps 10;

Qy 12 DVNNGIDGH--ALADRIGLDEAIEAWRLSFTGIDDDTWAALAAEQPLFEA-----TAD 62

AC 064476;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative DNA2-NAM7 helicase family protein.
GN AT2G19120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana.";
RL Nature 402:761-768 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC002392; AAD12029.1; -;
DR PIR: T00533; T00533.
DR GO: GO:0004386; F:helicase activity; IEA.
KW Helicase.
SQ SEQUENCE 1090 AA; 121522 MW; EB170342E18DCA1F CRC64;

Query Match 10.1%; Score 94; DB 10; Length 1090;
Best Local Similarity 26.9%; Pred. No. 31;
Matches 39; Conservative 17; Mismatches 55; Indels 34; Gaps 5;

QY 55 PLFE---ATADALVTFYDHLSEYERTQDLFA-----NSTKTVEQLKETQAEYLLGL 103
DB 441 PSYEQTPATPCFTTSFVHLHRSFNAPQLAAIHWAAMHTAAGTSSGVKK-QEWPFTL 499
QY 104 GRGEYDTYAAQRARIGKHIVLGGPDVYLGYATRYTGLLDALA-----D 150
DB 500 VQGPPGT-----GKTHTWGMLNVIHLVQYQYVTSLLKLAPEYTNQANECSSSD 550
QY 151 DVVADRGEEAAAADVELVARELPLM 175
DB 551 NILSGSIDEVLQNMHDNLFRLPKL 575

RESULT 15
Q8FUX2 PRELIMINARY; PRT; 514 AA.
AC Q8FUX2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Peptide ABC transporter, periplasmic peptide-binding protein.
GN BR1090.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,

RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
animal and plant pathogens and symbionts.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).
DR EMBL: AE014601; AAN34255.1; -;
DR TIGR: BR1090; -;
DR GO: GO:0005215; F:transporter activity; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR000914; SBP_bac_5.
DR Pfam: PF00496; SBP_bac_5; 2.
DR PROSITE: PS01040; SBP_BACTERIAL_5; 1.
KW Complete proteome.
SQ SEQUENCE 514 AA; 57239 MW; DB29DBD3B0AA8F8F CRC64;

Query Match 10.0%; Score 93.5; DB 16; Length 514;
Best Local Similarity 27.1%; Pred. No. 12;
Matches 49; Conservative 23; Mismatches 70; Indels 39; Gaps 8;

QY 1 MSNDNDTIVTADVRNGIDGHADRLGDLDEAELARLSFTGIDDDTMAALAAEQPLFEAT 60
DB 83 LSEDSKTL-TFKLRKGVKFH-----DGSMDKSDVDVLASFERYAKVS 122
QY 61 ADALVTFDYDHLSEYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTY--AAQAR 118
DB 123 PNAKVLDIVDH---YETPDDY-----TFVHLKEVNAAFDLTKSPVYFPFSIIPAEQKDK 174
QY 119 IGKIHVLGLGPDVYLGYATRYTGLLDALADDVVADRGEAA-----AAVDELVARP 171
DB 175 PARELDIITGTF-FKLGEWKRDShLYLEKFA-DYVADKGPASGVAGEKKVYVDKRVNF 232
QY 172 L 172
DB 233 L 233

Search completed: August 10, 2004, 15:28:42
Job time : 24.0533 secs

B
1011K